Journal of Speech and Hearing Research

June 1961 VOLUME 4 • NUMBER 2

Measurement of Auditory Threshold with a Special Purpose Analog Computer 105
EDGAR L. LOWELL, CAROL TROFFER WILLIAMS, ROBERT M. BALLINGER, AND DELPHI P. ALVIG

Intelligibility of Words Varying in Familiarity 113
ELMER OWENS

Intrinsic Cues and Consonant Perception 130
WILLIAM S-Y. WANG AND CHARLES J. FILLMORE

Binaural Hearing Aids and Speech Intelligibility 137
JAMES JERGER, RAYMOND CARHART, AND DONALD DIRKS

Hearing in Children: Acoustic Environment and Audiometer Performance 149
ELDON L. EAGLES AND LEO G. DOERFLER

Intensive Differential Sensitivity at Masked Threshold 164
ARNOLD M. SMALL, JR., AND FRED D. MINIFIE

Visual Paired-Associates Task with Deaf and Hearing Children 172
HANS G. FURTH

Lipreading and Letter Prediction 178
CORINNE M. TATOUL AND G. DON DAVIDSON

Shifts in Loudness of Pure Tones Associated with Contralateral Noise Stimulation 182
WILLIAM F. PRATHER

Letters to the Editor 194
ARTHUR S. HOUSE; JAMES F. CURTIS AND JAMES C. HARDY

Research News Notes 112, 171, 177, 193

American Speech and Hearing Association

OFFICERS

President
G. Paul Moore, Ph.D.
Northwestern University

Executive Vice-President
Jack Matthews, Ph.D.
University of Pittsburgh

Vice-President
Duane C. Spriestersbach, Ph.D.
University of Iowa

OFFICERS-ELECT

President-Elect
James F. Curtis, Ph.D.
University of Iowa

Vice-President-Elect
Elise S. Hahn, Ph.D.
University of California
Los Angeles

COUNCIL

The Officers and the following Councilors:

Stanley Ainsworth, Ph.D. (1961)
Oliver Bloodstein, Ph.D. (1960-62)
William G. Hardy, Ph.D. (1960-62)
Ruth B. Irwin, Ph.D. (1960-63)
James Jerger, Ph.D. (1959-61)
Wendell Johnson, Ph.D. (1959-62)
Hayes A. Newby, Ph.D. (1960-63)
Mildred Templin, Ph.D. (1961-63)
Joseph M. Wepman, Ph.D. (1961-63)
Dean E. Williams, Ph.D. (1959-61)

EXECUTIVE SECRETARY

Kenneth O. Johnson, Ph.D.

Journal of Speech and Hearing Research

EDITOR

Dorothy Sherman, Ph.D.

ASSISTANT TO THE EDITOR

Dorothy W. Moeller

STATISTICAL CONSULTANT

Leonard S. Feldt, Ph.D.

ASSOCIATE EDITORS

Oliver Bloodstein, Ph.D.
Arthur S. House, Ph.D.
James Jerger, Ph.D.
D. E. Morley, Ph.D.
Hildred Schuell, Ph.D.
Arnold M. Small, Ph.D.
William R. Tiffany, Ph.D.
John C. Webster, Ph.D.
Joseph M. Wepman, Ph.D.
Herbert N. Wright, Ph.D.

ASSISTANT EDITORS

Kenneth L. Moll, Ph.D.
Martin A. Young, Ph.D.

SPECIAL EDITORS

Ernest H. Henrikson, Ph.D.
Book Reviews

Martin F. Palmer, Sc.D.
Records

Elizabeth Moodie Prather, Ph.D.
Interlingua

BUSINESS MANAGER

Kenneth O. Johnson, Ph.D.

ASHA PUBLICATIONS BOARD

Virgil A. Anderson (1959-61); John Black (1961-63); Ernest J. Burgi, Representative, House of Delegates; Raymond Carhart (1960-62); Frederic Darley, Editor, *Monographs* (1960-62); Mary Huber, Editor, *Journal of Speech and Hearing Disorders* (1959-62); Kenneth O. Johnson, Business Manager of Publications and Editor, *Asha;* Dorothy Sherman, Editor, *Journal of Speech and Hearing Research* (1959-62); and Wendell Johnson, Chairman (1959-62).

Measurement of Auditory Threshold with a Special Purpose Analog Computer

EDGAR L. LOWELL

CAROL TROFFER WILLIAMS

ROBERT M. BALLINGER

DELPHI P. ALVIG

In a previous publication, Lowell and others (*10*) described a special purpose analog computer, Vannus I, designed to detect electrophysiological responses in the auditory system. The present article describes the use of the computer in the measurement of auditory thresholds.

Vannus I was designed to average evoked responses obtained by gross surface electrode techniques from the scalp of human subjects in response to auditory stimulation. The evoked responses are not visually detectable in the electroencephalograph (EEG) record because of the greater voltage of the spontaneous activity. Because the evoked responses have a relatively constant latency following auditory stimulation, they can be separated from the spontaneous activity by repeated sampling of the EEG output at fixed intervals following stimulation. The procedure rests on the assumption that the spontaneous activity is not temporally related to the stimulus and will average to zero with repeated sampling.

The process is viewed as similar to winnowing wheat from chaff, and the computer is named Vannus from the Latin term for the fan used in winnowing. In order to differentiate this procedure from the more elaborate digital methods suggested by Clark (*5*), or the photographic technique described by Peimer (*11*), or by Suzuki and Asawa (*12*), this process is referred to as evannation (a winnowing out) or, more properly, because of the time-locked property of the evoked response, temporal evannation.

This approach was suggested by an article by Geisler, Frishkopf, and Rosenblith (*9*) and the work of the Communications Biophysics Group of

Edgar L. Lowell (Ph.D., Harvard, 1952) is Administrator, John Tracy Clinic. Carol Troffer Williams (M.A., University of Washington, 1959) was at John Tracy Clinic at the time this study was made, now lives at 44 rue Sarrette, Paris. Robert M. Ballinger, John Tracy Clinic, and Delphi P. Alvig (M.A., University of Washington, 1961) are Research Assistants, John Tracy Clinic. This investigation was supported by a Research Grant from the Bureau of Crippled Children Services, Department of Public Health, State of California, and Children's Bureau, U. S. Department of Health, Education, and Welfare. Funds for the construction of Vannus I computer were made available through the generosity of North American Aviation Employees' 'Donate Once' Club, and the Western Foundation for Hearing Research.

the Research Laboratory of Electronics of the Massachusetts Institute of Technology (*6*) who had previously explored the use of digital computers for the measurement of evoked potentials. Prior to this, similar procedures had been suggested in 1954 by Dawson (7) who studied an electromechanical device for detection of small evoked potentials elicited by touch, and by the work of Barlow (*1*) and Brazier and Barlow (*3*) who have described a more sophisticated technique for performing a similar averaging process.

Equipment

Information is fed to the computer from an Offner Model T Electroencephalograph. The computer is equipped with a stimulus control unit which presents monaural stimulation in the form of a 10 μsec pulse from a Tektronix pulse generator through an attenuation circuit to a Permoflux PMR 16 earphone. The input is so brief that the acoustic output is a damped sinusoid whose frequency corresponds to the resonant frequency of the earphone. The evaluation of and communication about the intensity of click stimuli are complex. To simplify this problem, Troffer (*13*), in a separate psychophysical experiment, has compared the bursts of clicks with an 8 000 cps pure tone. References to intensity hereafter will refer to perceived loudness as compared with the audiometric tone.

The computer integrates only the EEG information passing through five preset gates. These gates are 2 msec in width and are set at 5 msec intervals starting 20 msec after the stimulus. The voltage passing through the gates is integrated and stored for 1 024 stimulations or approximately 1.75 min. At the completion of the run, the results from each of the five integrators are printed out on an Esterline Angus recorder. The capacity of the computer has been increased from that reported in the earlier presentation (*10*) by the addition of two more gates and integrators.

The computer can be programmed for a considerable range of variation in stimulus repetition rate, stimulus intensity and stimulus duration, gate width, position of gates following stimulus, and sample size.

Method

Subjects were 54 adults with normal hearing. They were seated in a dental chair with a headrest. EEG readings were taken with bipolar recordings between the left temporal and left occipital regions with a ground electrode at the frontal midline. Metal surface electrodes were held in place by an elastic strap. Electrode paste was applied under the electrodes. Only one channel of the EEG was studied. A Permoflux head set was used but only the right earphone was activated in these studies. Subjects were given reading material and requested to read during the testing. All tests were conducted in an acoustically treated section of the research laboratory of the John Tracy Clinic.

Test Runs. A test run consisted of 1 024 clicks presented at 100 msec intervals. Before each experimental session the subject's subjective threshold for clicks was determined by requiring him to press a button to indicate that he had heard a 2 sec burst of

clicks presented at varying intensities. The lowest intensity indicated in this fashion was designated as the subjective click threshold (SCT).

Experimental Sessions. An experimental session consisted of the presentation of a series of test runs. For each subject six test runs were presented at intensities above the SCT and four presented at intensities below the SCT. Four silent control periods, using the same gating and averaging procedure but without introducing a click through the earphones, were also employed.

Half of the experimental sessions started at 30 db re SCT as determined in a separate psychophysical experiment (*13*), and each successive run was presented at 5 db less intensity. Silent control periods were introduced on the fourth, eighth, twelfth, and fifteenth runs. The other half of the experimental sessions were all presented in the opposite or ascending order.

In all, there were 54 separate experimental sessions with 324 test runs above SCT, 54 at threshold, 216 below SCT, and 216 silent control periods. Some subjects participated in more than one experimental session.

In the interval between the runs, while the results were being printed out and the click intensity adjusted, radio music, adjusted to a comfortable listening level, was presented monaurally to the right ear.

Results

The first problem was to establish a criterion for the presence of an evoked response. The waveform under study, as suggested by animal experimentation (*2*, pp. 448-449), the work of the Massachusetts Institute of Technology Communications Biophysics Group (*6*), and the depth electrographic recordings from the human brain in response to clicks reported by Chatrian and others (*4*), is of sinusoidal form with a positive peak appearing around 20 msec, a negative peak at 30 to 35 msec, and a second positive peak at 45 to 50 msec. Individual subjects appear to vary around these points. Since the Vannus computer has only five gates, it was necessary to make certain arbitrary decisions concerning the parameters of the waveform that could be most efficiently studied with this equipment. Examination of preliminary data (*10*) suggested that the best correlation with subjective awareness and differentiation between suprathreshold and subthreshold or silent control responses is obtained by accepting as 'positive' only those average responses which show a negative-going change of at least 10 arbitrary voltage units between 25 and 30 msec, and a positive change of any

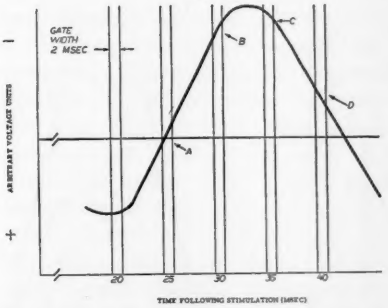

FIGURE 1. Section of hypothetical average evoked response illustrating response criterion which requires a 10 volt change between points A and B and a positive change between points C and D to qualify as a positive response.

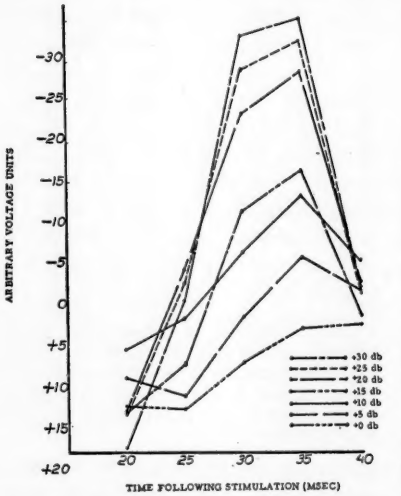

FIGURE 2. Average evoked response patterns at threshold (re subjective click threshold) and above.

amplitude between 35 and 40 msec. This response criterion is illustrated in Figure 1 in which a hypothetical average response is superimposed on the five gates. A response would be considered positive only when there was a negative change of at least 10 voltage units between points A and B

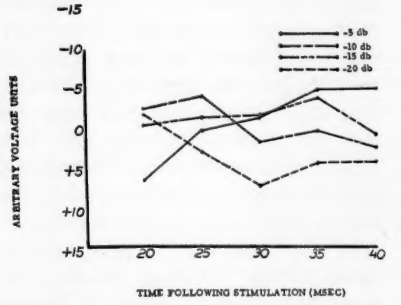

FIGURE 3. Average evoked response pattern at four subthreshold (re subjective click threshold) intensity levels.

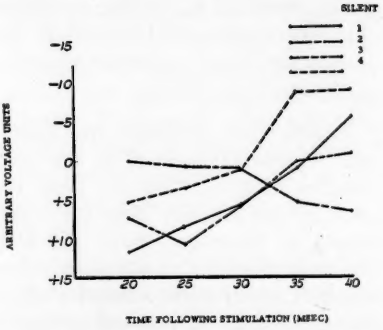

FIGURE 4. Average evoked response pattern for four silent control periods.

on Figure 1 and a positive change between points C and D. The voltage units are arbitrary, being a function of the amplification characteristics of the computer, the readout mechanism, and the EEG gain. They are, however, comparable between subjects. The response criterion is stated in terms of changes within an individual pattern, rather than with reference to the zero line on the amplitude scale, because individual patterns shift considerably around the zero line from run to run.

When the actual latency is shorter or longer, this criterion will not yield the most efficient information. These requirements seem justifiable, however, at the present level of knowledge.

The evoked response patterns for all 54 subjects are presented graphically in Figures 2 to 4: Figure 2, for those at threshold and above; Figure 3, at four subthreshold intensity levels; Figure 4, for four silent control periods. In these figures an upward deflection indicates that the occipital electrode was negative with respect to the temporal electrode. It will be noted that

TABLE 1. Percent of positive computer responses at different stimulus intensity levels for descending and ascending series, and for four silent control periods. (Total N = 54 experimental runs consisting of 1 024 clicks each at each intensity level.)

Stimulus Conditions	*Percent Positive Responses*		
	Descending	*Ascending*	*Total*
Intensity in db			
+30	93	70	82
+25	70	67	69
+20	74	70	72
+15	63	56	59
+10	41	44	43
+ 5	30	19	24
0	4	26	15
— 5	22	26	24
—10	7	19	13
—15	11	22	17
—20	19	4	11
Control Periods			
1	15	11	13
2	7	19	13
3	15	15	15
4	11	22	17

the polarities described herein are reversed from the earlier presentation (*10*). At that time it was assumed that the temporal electrode was the active electrode and the occipital electrode was the reference. A study by Geisler (*8*) has indicated that the magnitude of the average response is greatest from the occipital region. In order to standardize presentation of results, the occipital electrode is now considered the active electrode and the temporal electrode the reference.

The agreement between subjective awareness and positive responses as defined above is shown in Table 1. The agreement for all suprathreshold trials is 54% and increases in a fairly linear fashion with stimulus intensity ranging from 24% at +5 db to 82% at +30 db. The amount of error in the system can be estimated from the four subthreshold runs which averaged 16% responses, meeting the positive response criterion. This agreed rather well with the four silent control runs which averaged 14% positive responses.

It is well known in audiometric testing that descending series tend to yield lower thresholds than ascending series. The same results are shown in this study as illustrated by suprathreshold levels given in Table 1.

The individual errors were not equally distributed: 25 subjects or 44% of the runs accounted for 65% of the suprathreshold errors.

Discussion

As may be noted in Figure 1, the amplitude of the averaged evoked waveform increases systematically with the intensity of the stimulus. The percent of positive responses, however, also increases with intensity (Table 1). The reader may question whether the increase in amplitude of the average waveform is due primarily to the increasing number of positive responses. There is evidence that this is not the case. An examination of the average amplitude of only the *positive* responses, measured between the gate set at 25 msec and the gate set at 35 msec, shows the following results: 5 db above threshold, 26 volts; 10 db, 24 volts; 15 db, 30 volts; 20 db, 36 volts; 25 db, 38 volts; and 30 db, 42 volts. Thus, when any possible effect from the error responses is eliminated, the amplitude of the positive evoked responses still increases with intensity in a systematic fashion.

This preliminary study of a new technique of measuring auditory sensi-

tivity does not deal adequately with the statistical concept of threshold. It is apparent from an examination of Figure 1 that the average run at +5 and +10 db re SCT does not reach the 10 volt negative change between 25 and 30 msec required to meet the criterion for a positive response. In view of the fact that subjects were not 'attending' to the stimulus but were encouraged to read, the failure of the over-all averages at the two lowest intensities to meet the positive response criterion is not remarkable. It might be explained that this procedure was followed in preparation for the intended use of the computer with preschool deaf and hard of hearing children, where attention to the stimulus may be more difficult to achieve. If the results at +5 and +10 db re SCT are excluded from computation, the proportion of positive responses increases to 70%.

Some further information on the relation of the subject's attention to the threshold response may be inferred from a separate study in which the background electrical activity of the cortex was studied during the tests. The EEG records were analyzed in a systematic fashion for the percent of alpha rhythm present during the test period. Utilizing a click at 55 db above the SCT, and with the subject's eyes closed, it was found that those subjects showing less than 50% alpha in the EEG record gave significantly more positive responses than those with more than 50% alpha. Those in the group with 76% to 100% alpha were the lowest, showing only 52% positive responses. While many different explanations are possible, if an alert attentive state is associated with a desynchronized EEG pattern as contrasted with the relaxed state when alpha rhythm predominates, it might be inferred that those subjects who were more relaxed as indicated by the alpha rhythm were not as receptive to any incoming stimulation.

Even more evidence on the role of attention and the relation of positive computer responses to subjective awareness can be inferred from still another study in which the adult normal hearing subjects were requested to press a button to indicate when they heard the clicks. This procedure was initiated primarily to introduce some experimental control of the cortical activity by requiring attention.

While the percent of positive computer responses did not change significantly, the relation between the computer response and subjective awareness was clarified. With 20 additional subjects run under conditions similar to the main body of this study, but with the additional button pressing requirement, 92% of the 120 suprathreshold runs were correctly indicated by button pressing. Of the 20 click series at threshold, 75% were reported as heard; 10% of the 80 runs below threshold elicited button pressing, although for four of these the button was pressed during the run rather than at the beginning of the run as was the case with the suprathreshold clicks; 4% of the silent control runs elicited button pressing, and all occurred during the middle of the timed period during which averaging was taking place. In no case did the button pressing errors on the subthreshold or silent control runs agree with the posi-

tive computer responses which were scored as errors during these runs.

The error rate is influenced by the short sampling period of each test. The length of the experimental runs represents a compromise between the infinitely large sample ideally required by averaging theory and the practical considerations of the experimental test situation. The percent of positive responses can be increased and the percent of errors reduced by increasing the sample size. This may be accomplished by increasing the length of a single run, or by averaging together the results of several shorter runs. The latter method is preferable because it samples cortical activity at different periods and is less influenced by transitory interfering cortical activity. In the present study, the error rate for silent control runs 1, 2, and 3, which individually show 13%, 13%, and 15% error, is reduced to 7% error when the three silent runs for each experimental session are averaged together.

Summary

This paper describes a special purpose analog computer (Vannus I) designed to study evoked auditory responses by gross electrode techniques from the skulls of intact humans.

A study of 54 experimental runs with normal hearing adults is reported. Each subject was exposed to 1 024 clicks at 100 msec intervals at six suprathreshold and four subthreshold levels. Silent control runs without auditory stimulation also were employed.

Total responses showed a systematic change in the response form with changes in intensity. Total subthreshold and silent control runs were relatively flat and did not conform to the positive response criterion.

Individual response patterns showed 54% positive responses, ranging from 24% positive at +5 db to 82% positive at +30 db. There was good agreement between the percent of positive responses on the subthreshold and silent control runs.

Acknowledgment

The writers are indebted to Professor W. A. Rosenblith and Mr. C. D. Geisler of the Massachusetts Institute of Technology for introducing them to the ideas presented herein. They are also grateful to Dr. George Misrahy and Dr. Bruce Deatherage for helpful suggestions during the conduct of this investigation.

References

1. Barlow, J. S., An electronic method for detecting evoked responses of the brain and for reproducing their average waveforms. *Electroenceph. clin. Neurophysiol.*, 9, 1957, 340-343.
2. Brazier, Mary A. B., (Ed.), *Third Conference on Central Nervous System and Behavior*. New York: Josiah Macy, Jr., Fdn., 1960.
3. Brazier, Mary A. B., and Barlow, J. S., Some applications of correlation analysis to clinical problems in electroencephalography. *Electroenceph. clin. Neurophysiol.*, 8, 1956, 325-331.
4. Chatrian, G. E., Petersen, M. C., and Lazarte, J. A., Responses to clicks from the human brain: some depth electrographic observations. *Electroenceph. clin. Neurophysiol.*, 12, 1960, 479-489.
5. Clark, W. A., Jr., Average response computer (ARC-1). *Quart. Rep. Electronics Mass. Inst. Tech.*, April, 1958, 114-117.

6. Communications Biophysics Group of Research Laboratory of Electronics and Siebert, W., Processing neuroelectric data. *Tech. Rep. 351 Electronics Mass. Inst. Tech.*, July 1959.
7. Dawson, G. D., A summation technique for the detection of small evoked potentials. *Electroenceph. clin. Neurophysiol.*, 6, 1954, 65-84.
8. Geisler, C. D., Electrical responses to acoustic clicks recorded from human scalp. *Quart. Rep. Electronics Mass. Inst. Tech.*, April 1960, 137-143.
9. Geisler, C. D., Frishkopf, L. S., and Rosenblith, W. A., Extracranial responses to acoustic clicks in man. *Science*, 128, 1958, 1210-1211.
10. Lowell, E. L., Troffer, Carol I., Warburton, E. A., and Rushford, Georgina M., Temporal evannation: a new approach in diagnostic audiology. *J. Speech Hearing Dis.*, 25, 1960, 340-345.
11. Peimer, I. A., Local electrical responses in the cerebral cortex of man and their relationship to generalized reactions in the process of conditioned reflex activity. *Sechenov J. Physiol.*, 44, 9, 1958, 829-838.
12. Suzuki, T., and Asawa, I., Evoked potential of waking human brain to acoustic stimuli, a clinical study on its application to objective audiometry. *Acta Otolaryng.*, 48, 1957, 508-515.
13. Troffer, Carol I., Psychophysical matching of click bursts and pure tones. *John Tracy Clinic Tech. Memo. #8*, February 1960.

► *RESEARCH NEWS NOTE*

Infant Vocal Learning

► A study of infant vocal learning is being conducted at the University of Kansas Medical Center under the auspices of the Bureau of Child Research and the Children's Rehabilitation Unit. The project is directed toward testing apparatus and selecting rewards that are most effective in increasing infant vocalization. The learning paradigm of instrumental reward conditioning rather than that of classical conditioning is being employed because (a) the eliciting conditions for early infant babbling are not well known, and (b) spontaneous noisemaking is a previously ascertained characteristic feature of infant development. Furthermore, a functional analysis of verbal behavior satisfies the characteristics of operant rather than respondent activity.

The primary instrumentation, consisting of a sensitive voice-operated relay, has been built by Mr. Wesley Heisey of the University of Kansas Medical Center. The relay is actuated at approximately 55 db at a distance of two to three feet and then operates an external output (any of several reinforcers, such as a light, sound, or rocking cradle). Grason-Stadler equipment has been purchased to provide for accurate experimental duration timing, counting of responses, partial reinforcement training, and discriminative stimulus training.

It is hoped that the application of modern learning theory to infant speech will provide information heretofore not obtained from descriptive studies. This investigation is supported by PHS research grant M-3987 from the National Institute of Health.

Harris Winitz, Ph.D.
Project Director
Bureau of Child Research
University of Kansas
Lawrence, Kansas

Intelligibility of Words Varying in Familiarity

ELMER OWENS

Intelligibility lists were first used by communication engineers, notably at Bell Telephone Laboratories, to rate sound transmission systems. Typically, a speech sample was presented by a speaker through a transmission system to trained listeners who wrote or repeated what they heard. The percentages of correct responses served as efficiency ratings of the system. Other intelligibility lists, developed at the Harvard University Psycho-Acoustic Laboratories, found use in the clinical testing of hearing, the difference in procedure being simply that the listener became the variable while the speaker and transmission system were controlled. The listener's percentage of correct responses was taken as a measure of his ability to discriminate, or hear clearly, the sounds of English speech.

The Harvard lists, numbering 20, each contained 50 monosyllabic words. In selecting the words for each list, an attempt was made to achieve a phonetic composition representative of that found in everyday speech; hence the term PB (phonetically balanced) lists. An account of the Harvard PB-50 intelligibility lists is given by Egan (*4*).

Another series of intelligibility lists was developed at the Central Institute for the Deaf a few years later by Hirsh and others (*8*). These CID W-22 lists also comprised 50 monosyllabic words each. One of the differences between the CID W-22 lists and the Harvard PB-50s was that the former were composed of words selected on the basis of having greater familiarity; according to the authors, the purpose was to minimize the effect of differences in educational background among the persons tested.

After the CID W-22 lists had been tape recorded in final form, disc recordings were made of them, and the discrimination scores of normal hearing listeners on these recordings were compared with the scores of the same listeners on the 12 in., 33 1/3 rpm Technisonic Studio disc recordings of selected PB-50 lists. Scores were found to be consistently better on the W-22 recordings than on the PB-50 recordings at any given intensity level. Several explanations have been advanced; for example, Silverman and Hirsh (*16*) attached importance to the relatively short duration of the individual test words on the PB-50 recordings, as emitted by the speaker, Rush Hughes. Later in their discussion, however, these authors referred to an 'unknown fea-

Elmer Owens (Ph.D., University of Pittsburgh, 1959) is Assistant Professor of Audiology and Speech, Division of Otorhinolaryngology, University of California Medical Center. This article is based on a Ph.D. dissertation completed under the direction of Dr. Aubrey Epstein.

ture' in the Rush Hughes recordings making them more difficult.

Pertinent to the interests of the present study, little or no consideration has been given to the possibility that the better scores on the W-22 lists were in part directly related to the greater familiarity of the words. No doubt the possible effects of differences in word familiarity were overlooked because of the traditional emphasis on the phonetic make-up of the words constituting the test lists. Apparently the assumption was that if a list of test words contained all the phonetic elements in proportion to their occurrence in everyday speech, there would be a one-to-one relationship between a discrimination score and hearing acuity in everyday communication; that is, the relationship would depend entirely upon the hearing acuity for individual sounds making up the test words and would not be influenced to any great extent by other word attributes such as familiarity.

At the present time, both series of lists are used in various audiology

TABLE 1. Number of words from the PB-50 and the W-22 intelligibility lists in each of four familiarity categories. Categories are based on the number of times a word appears in the Lorge count of 4 500 000 words: unfamiliar, 0 to 100 times in the 4 500 000 words; moderately familiar, 100 to 1 000; highly familiar, 1 000 to 10 000; extremely familiar, 10 000 to 100 000.

	List	*Familiarity Categories*				*Total*
		Non	*Moderate*	*High*	*Extreme*	*Words*
PB-50	1	21	18	5	6	50
	2	26	15	8	1	50
	3	19	23	8	0	50
	4	23	19	6	2	50
	5	19	23	8	0	50
	6	32	11	6	1	50
	7	25	18	4	3	50
	8	20	19	9	2	50
	9	21	22	6	1	50
	10	26	16	8	0	50
	11	23	20	5	2	50
	12	23	19	5	3	50
	13	21	22	5	2	50
	14	20	22	6	2	50
	15	24	16	8	2	50
	16	25	18	6	1	50
	17	18	24	5	3	50
	18	24	23	2	1	50
	19	20	21	7	2	50
	20	25	19	4	2	50
	Total	455	388	121	36	
	Mean	22.8	19.4	6.0	1.8	
W-22	1	8	18	11	13	50
	2	5	23	15	7	50
	3	4	20	15	11	50
	4	3	18	18	11	50
	Total	20	79	59	42	
	Mean	5	19.7	14.7	10.5	

centers. It is the general impression of most clinicians that the PB-50s are more difficult than the W-22s, whether the lists are presented by live voice or recordings. Moreover, within each series, but particularly within the PB-50s, some of the lists appear more difficult than others. Corroborative research on the latter observation has been reported by Eldert and Davis (*5*).

Research investigations in the area of visual and auditory perception, largely within the past eight or 10 years, appeared to have direct bearing on the relative difficulty of intelligibility lists. These studies, concerned with the effects of word familiarity upon intelligibility, employed an operational definition of familiarity: a word was more or less familiar according to its frequency of usage in a language. The frequency of usage in turn was obtained from one or a combination of the few English word counts generally considered valid, such as the Lorge (*18*), the Lorge-Thorndike (*18*), and the Dewey (*2*). Some of the experimenters, on the other hand, controlled frequency of usage directly by means of practice sessions introduced as part of the experimental design. Findings of the earlier studies (*11, 12, 17*), dealing with visual perception, were consistent in that, when exposed tachistoscopically, words of greater familiarity were significantly easier for subjects to recognize than words of lesser familiarity. Results of later studies, concerned primarily with auditory perception (*1, 10, 14, 15*), were also consistent: test words presented under various conditions of masking noise were easier to identify insofar as they were more familiar to the subjects.

On the basis of these studies and the apparent differences in difficulty between the PB-50 and W-22 lists as used clinically, an inspection of the words on the lists was made relative to their familiarity. Frequencies of usage were obtained from the Lorge count (*18*) and tabulated as in Table 1. All words in both series of lists were categorized under four arbitrarily selected usage-frequency groupings: 0 to 100; 100 to 1 000; 1 000 to 10 000; and 10 000 to 100 000. Any word placed in the first category occurred less than 100 times in the 4 500 000 words constituting the Lorge count; any word placed in the second category occurred between 100 and 1 000 times in the 4 500 000; and so on. Corresponding familiarity classifications chosen were 'unfamiliar,' 'moderately familiar,' 'highly familiar,' and 'extremely familiar.' Examination of Table 1 reveals the following: (a) the over-all familiarity of the words on the W-22 lists is markedly higher than that for the words on the PB-50 lists; (b) on all individual lists the range of familiarity is wide; (c) variations in over-all familiarity occur from list to list within each series, but are more noticeable among the PB-50 lists.

Further study of the effects of differences in word familiarity on intelligibility seemed indicated, the aim being directed primarily to the investigation of possible implications for discrimination testing. The hypothesis tested was as follows: when lists of monosyllabic words are presented once and without previous practice to normal hearing listeners under varying conditions of distortion induced by low-pass filtering, discrimination scores in terms of error responses will vary significantly from

TABLE 2. Words in devised intelligibility lists, each word chosen on the basis of frequency of usage in the 4 500 000 words of the Lorge count. The lists being compared (1A, 1B, and 1C; 2A and 2B; 3A and 3B) were varied systematically in terms of the Lorge count and were matched phonetically with respect to initial and final sounds; matching of vowel sounds was only approximate. Frequencies of usage for the first list in each case (that is, for lists 1A, 2A, and 3A) were 0 to 100. Frequencies of usage for the other lists were as follows: 1B, 150 to 500; 1C, 1 000 to 2 000; 2B, 3 649 to 9 499; 3B, 2 605 to 52 101.

Word Lists						
1A	*1B*	*1C*	*2A*	*2B*	*3A*	*3B*
kite	cat	cut	tame	time	shy	she
bug	beg	big	lurk	like	oat	it
skull	scale	school	ooze	is	hoe	he
bide	bird	bad	gout	get	heed	had
knit	net	note	meek	make	bat	but
fad	feed	food	cone	can	hose	his
wick	wake	work	jest	just	hive	have
shay	shoe	show	moan	man	woe	we
den	dine	done	dude	did	hum	him
furl	file	feel	cam	come	shale	shall*
lease	loose	less	goal	girl	ape	up
cape	cap	keep	dune	down	haze	has
fern	fun	fine	feast	first	thine	than
mole	meal	mile	goad	good	ace	us
roan	rain	run	tack	take	cud	could
booth	bath	both	loaf	life	etch	each
noun	noon	known	hiss	house	shad	should
loaf	leaf	laugh	tile	tell		
sane	sin	soon	plus	place		
boost	boast	best	ram	room		

*The word 'shall' is an exception, having a frequency count of only 708.

list to list depending upon the familiarity of the words composing the lists.

Procedure

Development of Word Lists. As in the studies mentioned earlier, an operational definition of familiarity was used in the selection of test words: a word is more or less familiar depending upon its frequency of usage in the English language as indicated by a generally accepted word count. Of several word counts considered, that by Lorge (*18*) was selected as clearly the most valuable for purposes of this study. This count included about 4 500 000 words from five of the most widely read magazines in the years 1927 through 1938: *Saturday Evening Post, Woman's Home Companion, Ladies' Home Journal, True Story,* and *Reader's Digest.* The Dewey (*2*) and the Lorge-Thorndike (*18*) counts, the former based upon spoken as well as written material and the latter on written material alone, were employed as checks; that is, the test words chosen from the Lorge count had to have a roughly equivalent usage-frequency rating on either or both the Dewey and Lorge-Thorndike counts.

The use of word counts based primarily upon written material might be

questioned in that auditory and not visual responses were being studied. Several observations may be cited in support of employing these counts: (a) Howes (*10*) reported a high correlation (.75) between the Lorge-Thorndike frequency count and words making up the oral usage of college students; (b) Postman and Rosenzweig (*14*) pointed out pronounced transfer effects on auditory recognition when previous practice, varying frequency of usage, had been restricted to words as visual stimuli; (c) Rosenzweig and Postman (*15*) stated that counts of written and spoken words in French show fairly close correspondence. In addition, it may be noted that the literature forming the basis for the Lorge count tends toward a conversational style.

Word lists developed for experimental manipulation are shown in Table 2. Words constituting the test lists were selected on the basis of their phonetic make-up and frequency of usage, the aim having been to vary the over-all familiarity of the lists being compared while holding constant their phonetic content. Because the exigencies of selecting test items in terms of both familiarity and phonetic content demanded a large supply of words at each of several frequency-of-usage levels, an arbitrary procedure was necessitated in choosing the levels, or bands, to serve as familiarity categories. Usage frequencies for all the PB-50 and W-22 words were obtained from the Lorge count and tabulated to see whether the words tended to cluster around certain points in the frequency-of-usage range. Several prominent clusters appeared, the first at frequencies of usage below 100; that is, a large number of words were found to occur less than 100 times in the 4 500 000 words of the Lorge count. Another cluster appeared at frequencies of usage between 200 and 400, and still another between 1 000 and 2 000. Noticeably fewer clusters were observed above the usage frequency of 2 000. Accordingly, the three usage-frequency bands where the clusters occurred were selected as categories for lists 1A, 1B, and 1C. As finally arranged, these three lists comprised words in usage-frequency categories of 0 to 100, 150 to 500, and 1 000 to 2 000. Familiarity classifications assigned were 'unfamiliar,' 'moderately familiar,' and 'highly familiar.'

In lists 1A-1B-1C, three-word combinations, such as 'bug-beg-big' or 'kite-cat-cut,' were employed, one word being assigned to each test list according to its frequency of usage. For example, 'bug,' with a usage frequency of 34, was assigned to list 1A; 'beg,' with a usage frequency of 302, was assigned to list 1B; and 'big,' with a usage frequency of 1 773, was assigned to list 1C. Thus, these three-word combinations were varied systematically with respect to frequency of usage and matched with respect to initial and final consonant sounds. Despite unexpected difficulty in finding such three-word combinations, only one notable exception was necessary in forming the lists: the word 'work' on list 1C was outside the 1 000 to 2 000 range, having a usage frequency of 2 720.

An attempt to equalize the vowel content of the lists did not meet with complete success, again owing to the scarcity of the above-described three-

word combinations. The matching of vowels, however, though highly desirable, was considered far less crucial than the matching of consonants because the filtering used as a distorting agent would have its major effect upon the consonants.

The remaining four lists were paired, 2A being matched in phonetic content and varied in familiarity with 2B, and 3A matched phonetically and varied in familiarity with 3B. List 2A was composed of words having a usage frequency of less than 100 on the Lorge count, while its counterpart, list 2B, comprised words of extremely high frequency of usage. List 2A was classified as 'unfamiliar'; list 2B, as 'extremely familiar.' List 3A also was composed of words under 100 on the Lorge count; its paired list, 3B, made up of words having high frequency of usage over a wide range, was distinctive in that it included only contextual, or interstitial, words. List 3A was classified as 'unfamiliar,' and list 3B as 'extremely familiar, contextual.'

Many words were omitted from the test lists on the grounds that they might confound or invalidate the experimental results. Among these were homophones, proper names, slang terms, words that might produce emotional or shock effects, and words unique to a specific vocation, sport, religion, or sex. (Some homophones were included on the condition that their combined frequency counts did not change the frequency category.)

It will be noted that all test lists contain 20 words except 3A and 3B, which contain 17 each; three pairs of words were eliminated from these two lists when closer examination revealed that they failed in one way or another to satisfy the selection criteria.

Finally, words were selected to cover as wide a variety of speech sounds as possible. No attempt was made to achieve phonetic balance within lists.

Recording and Filtering. The words of all seven lists were recorded on an Ampex 350 tape recorder by a male speaker of General American dialect using a high quality dynamic microphone. For each test item, the carrier phrase 'write the word' was monitored at the zero point on the VU meter and the test word emitted at the end of the carrier phrase with natural inflection and stress as in routine audiologic procedure. A 1 000 cps calibration tone was recorded on the tape at the same VU point zero in order to provide a reference level. The time pattern followed by the speaker allowed about five seconds for the subjects to write each response.

These seven lists recorded on the master tape were then filtered using an Allison Variable Filter, Model 2B, set at low pass with cut-off frequencies of 3 120, 2 040, 1 560, 960, 780, and 540 cps. The slope at the cut-off points was approximately 30 db per octave. It was thought that the 'hearing losses' resulting from the low-pass filtering at these points would approximate many of the hearing loss curves actually shown on audiograms, thus permitting a basis for generalization. Of course, the tenuousness of such generalization can be appreciated when it is realized that the frequency distortion caused by aural pathology cannot be reduced simply to a mechanical elimination or attenuation of frequencies and when consideration is given to other possible

concomitants of signal distortion in aural pathology.

In the filtering process, the recorded test words were fed from the output of the Ampex 350, through the Allison low-pass filter, to the input of a Berlant (Model 132) tape recorder. The filtered lists were recorded on seven individual 7.5 in. tape reels, one reel for each of the six filtering conditions and an additional one for an unfiltered condition. At the beginning of each of the tape reels, the 1 000 cps calibration tone was inserted and the input of the Berlant tape recorder adjusted to insure a constant level of recording for all reels. Throughout, the filter was manipulated in a way permitting a clear, unfiltered recording of the announcements of list and item numbers.

Test Administration. The unfiltered tape reel and the six prepared under the different filtering conditions were fed from the line output of the Berlant tape recorder to a Hewlett-Packard attenuator, Model 350 B, and thence through a matching transformer and a high fidelity amplifier to eight high quality speaker systems positioned two in each corner of a sound-treated listening room large enough for as many as 25 listeners seated around three tables used for writing purposes. The multiple speaker systems suitably loaded the listening room so that, in essence, some of the advantages of earphone listening were achieved while utilizing the superior high frequency response of correctly baffled loud speakers (± 5 db, 60 to 16 000 cps). The control room, which housed the Berlant tape recorder, was separated from the listening room acoustically and otherwise, but a window between the two rooms permitted visual contact between listeners and experimenter.

Listeners, or subjects, were college students enrolled in speech classes, each class numbering 10 to 20. Age and sex distribution were approximately equivalent from class to class. Each class listened as a group to all seven lists under one filtering condition, the condition for each group having been selected largely at random, although departure from random selection was made occasionally in the interest of achieving a final N of 30 for each condition.

After each group had been seated in the listening room, the experimenter gave a short verbal description of the experiment for general orientation. Specific recorded instructions were presented at the beginning of each tape reel as follows:

> You are about to hear a series of words. First you will hear the number of the word to be given, then a few seconds later the phrase 'write the word,' followed by the test word itself. Write the test word in the appropriate blank space on your answer form. For example, if the speaker says, 'number one, write the word *house*,' you would simply write *house* after number one. Keep in mind that you need write only the test word, not the whole phrase. Some of the words may sound distorted. Be sure to fill in each blank with a word even though you must guess. It is most important that you fill in each blank with a word even though you may not be sure it is correct. So do not hesitate to guess. The first three items are for practice.

After the three practice items, the experimenter returned to the listening room to answer questions.

For each listening condition, the test words were presented at 50 db above the threshold of detectability for the specific list. All listening was done in

relative quiet, and the level of 50 db above the threshold of detectability afforded a comfortable loudness for all groups.

Each group was given a rest period of a minute or two at the end of the third list, but invariably the subjects considered the rest period neither necessary nor desirable. Their actions and comments indicated no evidence of fatigue effects in any of the listening conditions.

Immediately following the half-hour listening session, hearing tests were given. Two experienced audiometrists using standard clinical pure-tone audiometers that met the American Standards Association specifications did individual sweep checks at a 15 db sensation level, covering frequencies 250 through 8 000 cps at octave intervals.

The tape recordings of the seven word lists were presented to 17 groups in all, each group listening to the words under one of the seven filtered conditions. Subjects with a foreign dialect were eliminated as were those having a hearing loss, which was defined as failure on two or more screening frequencies in either ear or failure on the same frequency in both ears. The subjects remaining totaled 30 or more under each condition. Where more than 30 remained for any condition, the N of 30 was achieved by random elimination.

Results and Discussion

Scoring. Subjects were scored on each list according to the number of errors. The scoring procedures were essentially simple in that if the written response for each item was not the same as the test word, it was counted a mistake. Misspelled words were counted correct if they had the same initial and final sounds as the stimulus word, and if the letter combination for the vowel was among the typical spellings listed by Fairbanks (*6*). This arbitrary rule seemed an appropriate solution to the scoring for all but a few of the misspelled words.

Before direct analysis of the data, *t* tests were run on the mean over-all error scores of the groups under each condition. In the majority of listening conditions, there were two groups of listeners; in a few of the conditions, however, there were three groups. In each of the latter instances, the two smaller groups were combined and treated as one for purposes of the *t* test. None of the differences reached significance at the 5% level. Accordingly, there was no evidence that the two groups making up the total of 30 subjects under each listening condition were not from the same population. Further, the assumption was made that all seven groups of 30 were from the same population of college students enrolled in speech classes. A priori considerations would also lead to the same assumption, since age range, academic interest, and proportion of men and women were roughly equivalent for all groups. The experimenter felt that there were no factors present that would result in systematic differences with respect to performance on the experimental task. Thus, although two or three groups of subjects listened under each of the filtered conditions at different times, statistical tests as well as a priori considerations indicated that for purposes of statistical analysis there should be no objection to combining

TABLE 3. Summary of analyses of variance for error scores of seven groups of subjects tested under different low-pass filtered conditions (unfiltered; 3 120, 2 040, 1 560, 960, 780, and 540 cps).

Lists	*Source*	*df*		*ms*	*F**
1A, 1B, 1C,	Between Conditions	6		4 406.60	736.89†
	Between Subjects in Same Group	203		5.98	
	Total Between Subjects		209		
	Between Lists	2		1 174.40	602.25†
	Interaction: Lists x Conditions	12		71.64	36.73†
	Interaction: Pooled Subjects x Lists	406		1.95	
	Total Within Subjects		420		
2A, 2B	Between Conditions	6		2 747.16	756.79†
	Between Subjects in Same Group	203		3.63	
	Total Between Subjects		209		
	Between Lists	1		1 419.01	265.18†
	Interaction: Lists x Conditions	6		53.51	27.30†
	Interaction: Pooled Subjects x Lists	203		1.96	
	Total Within Subjects		210		
3A, 3B	Between Conditions	6		1 898.27	493.04†
	Between Subjects in Same Group	203		3.85	
	Total Between Subjects		209		
	Between Lists	1		242.31	118.20†
	Interaction: Lists x Conditions	6		10.52	5.13†
	Interaction: Pooled Subjects x Lists	203		2.05	
	Total Within Subjects		210		

**F* ratios: ms_C/ms_S; ms_L/ms_{SL}; ms_{LC}/ms_{SL}.
†Significant at the .001 level.

the groups and assuming an N of 30 for each condition. Furthermore, it seemed logical to assume that all seven groups of 30 were from the same population.

Analyses of Variance and Differences between Individual Pairs of Test Lists. Error scores on each list were totaled and treated statistically by three methods-by-trials analyses of variance for repeated measures after the manner suggested by Edwards (*3*). In the present study, the various filtered conditions corresponded to 'methods' and the test lists to 'trials.' One analysis of variance was accomplished for lists 1A (unfamiliar), 1B (moderately familiar), and 1C (highly familiar) in all conditions; a second for 2A (unfamiliar) and 2B (extremely familiar) in all conditions; and a third for 3A (unfamiliar) and 3B (extremely familiar, contextual) in all conditions. Three separate analyses were performed since the main interest was in comparing only those lists matched phonetically.

As seen in Table 3, all three *F* tests—between conditions, between lists, and interaction between lists and conditions—were significant beyond the .001 level.

TABLE 4. Evaluation by the critical difference technique after Lindquist (*13*) of the differences in mean error scores between individual pairs of test lists following an over-all analysis of variance.

Conditions	*Lists*				
	1A, 1B	*1A, 1C*	*1B, 1C*	*2A, 2B*	*3A, 3B*
Unfiltered	1.20	1.30	.10	.77	.74
3 120 cps LP	2.50*	2.50*	.00	2.13*	1.90*
2 040 cps LP	3.80*	5.50*	1.70*	3.67*	1.13
1 560 cps LP	5.57*	8.77*	3.20*	3.54*	2.30*
960 cps LP	2.80*	7.97*	5.17*	4.97*	2.84*
780 cps LP	.23	3.33*	3.10*	6.63*	.64
540 cps LP	2.33*	3.67*	1.34*	4.04*	1.10

*Significant at the .001 level.

(When examination of the raw data had revealed skewness in several distributions of error scores and heterogeneity of variance between some of the lists to be compared, it was realized that the assumptions underlying the analysis of variance were not met with exactitude. The stringent confidence level criterion of .001 was adopted as a compensatory measure.)

For the study of relationships between the individual pairs of means under each condition following the analyses of variance, the critical-difference technique as described by Lindquist (*13*) was employed. The results are shown in Table 4. Differences between the mean error scores of lists 1A (unfamiliar) and 1B (moderately familiar) were significant at the .001 level in all conditions except the unfiltered and the 780 cps low-pass, and the differences between the means of 1B (moderately familiar) and 1C (highly familiar) were significant at the .001 level in all but the unfiltered and 3 120 cps low-pass conditions. Differences between the means of 1A (unfamiliar) and 1C (highly familiar) were significant at the .001 level in all conditions except the unfiltered, as were the differences between the means of 2A (unfamiliar) and 2B (extremely familiar). The results for lists 3A (unfamiliar) and 3B (extremely familiar, contextual) indicated a significant difference (.001 level) in only the 3 120, 1 560, and 960 cps low-pass conditions. The direction of all the above differences indicated greater intelligibility for the more familiar lists.

The number of errors made in the unfiltered condition was surprising, particularly on the 1A (unfamiliar) list where they were so numerous as to be greater than the errors of lists 1B (moderately familiar) and 1C (highly familiar) at the .01 confidence level. The implication is that normal-hearing persons listening under the best of circumstances will make a substantial number of mistakes on any discrimination test that includes words of low familiarity. It should be mentioned that the words on the 'unfamiliar' lists were all bona fide in that there was no in-

tention of making them particularly difficult by selection. All words were known by a pilot group of nine listeners tested before the experiment proper; these listeners made perfect scores on the test lists as did many of the subjects in the experimental situation. Analysis indicated, also, that the errors were not restricted to a few of the unfamiliar words but were well distributed over all. List 1A seemed more difficult than the other two 'unfamiliar' lists (2A and 3A), probably because it contained more sibilant sounds. Words containing these sounds appeared more frequently in error throughout the experiment.

The lack of a significant difference between 3A (unfamiliar) and 3B (extremely familiar, contextual) in the 2 040, 780, and 540 cps low-pass conditions lent support to the observation of Howes (*10*) that contextual words do not appear to behave in the same way as other words relative to the intelligibility, frequency-of-usage relationship. Despite an extremely high frequency of occurrence according to the Lorge count, the intelligibility of these words seemed to break down under severe distortion to much the same extent as the intelligibility of the low-usage-frequency words with which they were matched phonetically. An inspection of the mean error scores in Table 5 indicates that even in the conditions where the differences between these lists were significant at the .001 level, the actual differences were noticeably less than between, say, lists 1A (unfamiliar) and 1C (highly familiar). Howes (*10*) theorizes that the explanation may lie in these words having importance only in context and being without meaning when they stand alone, as in an intelligibility list. Also, he feels that they tend to be skipped in reading, rendering their frequency counts spuriously high in terms of actual usage.

With a view to checking the possibility that the heterogeneity of variance between some of the test lists and the skewness of certain of the distributions

TABLE 5. Summary of means and standard deviations for distributions of errors on test lists under unfiltered and low-pass (LP) filtered conditions.

Conditions		*Lists*						
		1A	*1B*	*1C*	*2A*	*2B*	*3A*	*3B*
Unfiltered	Mean	1.50	.30	.20	.93	.16	.80	.06
	SD	1.14	.58	.40	.89	.37	1.07	.24
3 120 cps LP	Mean	3.46	.96	.96	2.66	.53	2.23	.33
	SD	1.83	1.04	1.04	1.13	.80	1.32	.47
2 040 cps LP	Mean	7.23	3.43	1.73	4.80	1.13	4.26	3.13
	SD	2.00	1.47	.89	1.64	1.60	2.04	1.20
1 560 cps LP	Mean	10.60	5.03	1.83	6.50	2.96	7.36	5.06
	SD	2.48	1.76	1.29	1.92	1.04	2.01	1.74
960 cps LP	Mean	15.80	13.00	7.83	12.30	7.33	11.20	8.36
	SD	2.37	2.46	1.89	2.42	2.07	2.11	2.45
780 cps LP	Mean	18.13	17.90	14.80	18.83	12.20	14.10	13.46
	SD	1.45	1.88	2.90	1.12	3.22	2.02	2.24
540 cps LP	Mean	19.23	16.90	15.56	19.20	15.16	14.53	13.43
	SD	.91	1.73	1.80	.94	2.31	1.11	1.78

TABLE 6. Differences in mean error scores between selected pairs of test lists under the 960 cps low-pass filtered condition.

Lists	*Means*	*Diff*	*SE*	*t*
1A (unfamiliar)	15.80			
1B (moderately familiar)	13.00	2.80	.34	8.24*
1A (unfamiliar)	15.80			
1C (highly familiar)	7.83	7.97	.45	17.71*
1B (moderately familiar)	13.00			
1C (highly familiar)	7.83	5.17	.42	12.28*
2A (unfamiliar)	12.30			
2B (extremely familiar)	7.33	4.97	.44	11.59*
3A (unfamiliar)	11.20			
3B (extremely familiar, contextual)	8.36	2.84	.39	7.76*

*Significant beyond the .001 level.

biased the analyses of variance beyond that compensated for by the adoption of the highly stringent significance level, individual *t* tests for related measures, independent of any analysis of variance, were accomplished for the mean error scores of the lists under the 960 cps low-pass condition. Examination of the raw data indicated that the errors on all word lists under this condition assumed an essentially normal distribution, and the variance appeared approximately equal from list to list. The *t*s obtained were for differences between the mean error scores of lists 1A and 1B (unfamiliar and moderately familiar), 1A and 1C (unfamiliar and highly familiar), 1B and 1C (moderately familiar and highly familiar), 2A and 2B (unfamiliar and extremely familiar), and 3A and 3B (unfamiliar and extremely familiar, contextual). All *t*s were significant beyond the .001 level in the direction of greater intelligibility for the more familiar lists (Table 6). The significant difference between lists 1A and 1B is of especial importance in demonstrating the effect of even small variations in familiarity.

The results of the *t* test under the 960 cps low-pass condition as well as the analyses of variance over all filtered conditions leave little doubt that when monosyllabic words were presented in this study after the manner commonly employed in discrimination testing—specifically, one presentation without previous practice—the discrimination ability of normal hearing listeners was markedly affected by word familiarity as defined by frequency of usage based on the Lorge count. Moreover, it seems safe to generalize these results to other populations: other things being equal, any individual would be expected to score better on a list containing a preponderance of highly familiar words than he would on a list in which the words were predominantly of low familiarity.

Relating the experimental findings to the PB-50 and W-22 lists, observations can be made as follows: (a) the marked difference in over-all familiarity be-

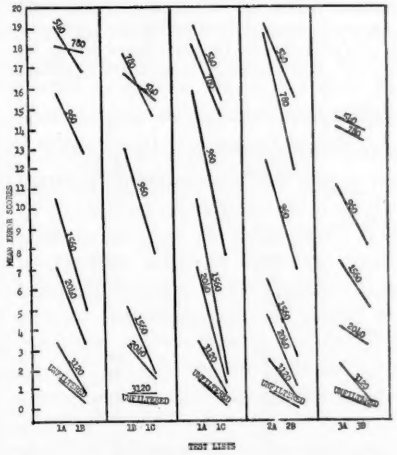

FIGURE 1. Comparisons between phonetically matched word lists in terms of error scores under different filtered conditions. Parameters are low-pass filter cut-off points in cps. All lists have 20 words with the exception of 3A and 3B, which have 17 each.

tween the two series, that of the W-22 being noticeably higher (see Table 1), would play an important role in making the W-22 lists easier to understand; (b) differences in over-all familiarity among word lists within each series, but particularly among the PB-50 lists (Table 1), would tend to cause variations in listener difficulty from list to list; (c) to the extent that a discrimination score is considered a measure of hearing acuity for the phonetic elements of speech, the wide range in word familiarity within each of the lists in both series would have confounding effects.

In refining present monosyllabic-word intelligibility lists or developing new ones, the familiarity of the words constituting the lists must be rigidly controlled if the purpose of the test is to measure the listener's ability to discriminate between speech sounds. Ideally, all test words should be within the narrowest possible usage-frequency band according to an acceptable word count. The selection of the frequency-of-usage band would depend upon the degree of difficulty desired. A practical working assumption might be that an intelligibility test should be difficult to the extent that it is sensitive to small differences in discrimination ability among listeners, but not so difficult as to cause any listener to become discouraged or otherwise lose motivation to do his best. Obviously, the difficulty of a test list could be controlled by selecting higher or lower frequency-of-usage bands for the test words while keeping constant the phonetic content.

Conditions. The *F*s for conditions in all three analyses of variance were significant beyond the .001 level. Figure 1 shows the increase in errors as more of the high frequencies were filtered out. The only reversal occurred in list 1B (moderately familiar), where more errors were found in the 780 cps low-pass condition than in the 540 cps low-pass. The increase in errors from condition to condition was not a steady phenomenon; noticeably greater differences occurred generally at the 960 cps and 780 cps low-pass points.

Although French and Steinberg (7) studied discrimination of nonsense syllables and employed a filter with an infinitely sharper cut-off than that used in the present study, they found similar trends in errors with respect to low-pass filter settings. Hirsh, Reynolds, and Joseph (*9*), using monosyllabic words and a filter cut-off rate of 60 db per octave, also found the same general

TABLE 7. Comparison of discrimination scores obtained on list 1A (unfamiliar words) with those of French and Steinberg (nonsense syllables) and of those obtained on list 2A (extremely familiar words) with those of Hirsh and others (W-22 words). Discrimination scores for French and Steinberg (7) and for Hirsh and others (*9*) were estimated from charts presented in their studies.

Cut-off Frequency (cps LP)	*Discrimination Scores*	*Cut-off Frequency (cps LP)*	*Discrimination Scores*
French and Steinberg		*1A*	
1 950	70%	2 040	64%
1 500	50%	1 560	47%
1 000	27%	960	21%
750	15%	780	9%
Hirsh and Others		*2A*	
2 000	90%	2 040	94%
1 500	80%	1 560	85%
1 000	63%	960	63%
700	50%	780	39%
500	20%	540	24%

trends. For purposes of comparison, approximations of discrimination scores at low-pass filter cut-off points corresponding to those of the present study were obtained from French and Steinberg (7) for nonsense syllables and from Hirsh, Reynolds, and Joseph (*9*) for monosyllables; at the same time, the error scores on lists 1A (unfamiliar) and 2B (extremely familiar) were converted to discrimination scores. As shown in Table 7, the scores for the French and Steinberg nonsense syllables and those for the 'unfamiliar' monosyllables of the present study correspond closely under the various filtered conditions. Similarly, the discrimination scores for Hirsh, Reynolds, and Joseph's monosyllabic words agree closely with scores for the 'extremely familiar' monosyllables of the present study. The latter authors used the W-22 words, which on the average are of fairly high familiarity (Table 1).

The continuous, albeit irregular, increase in discrimination errors on all lists as distortion becomes greater, points up the basic importance of the phonetic elements. As sounds are eliminated, attenuated, or distorted, all words are more difficult to discriminate. Thus, although word familiarity plays an important role in the discrimination of speech, phonetic content remains the dominant factor. This observation disagrees with the statement of Howes (*10*) that 'word frequency and word length together account for such a large proportion of the intelligibility variance that the intelligibility of a word—at least in the presence of a wide-spectrum noise—must be nearly independent of the relative intelligibilities of its constituent phonemes.'

In relating tenuously the filter settings of this study and high frequency hearing loss, it would appear that a person with such loss would have noticeably more difficulty discriminating speech sounds to the extent that fre-

quencies below 1 560 cps are involved. Also, the range of frequencies between 1 560 cps and 780 cps would be expected to have a disproportionate influence upon his discrimination ability.

Interactions. The *F*s for interaction between conditions and test lists in the three analyses of variance were all significant beyond the .001 level. The interactions between listening conditions and changes in intelligibility relative to familiarity can be judged by the slope of the lines in Figure 1. A perusal of the trends in differences between error means under the various conditions, depicted in Table 5, would also be informative regarding interaction effects. The magnitudes of change in intelligibility with differences in familiarity were generally greatest at the 1 560 and 960 cps cut-off points and somewhat less at the cut-off points of 2 040 and 780 cps.

Interaction effects for 3A (unfamiliar) and 3B (extremely familiar, contextual) were relatively small and invariable, offering further evidence that the contextual words do not behave the same as other words with respect to their familiarity-intelligibility relations.

Postman and Rosenzweig (*14*) obtained much the same interaction effects in their study of the intelligibility of words in various signal-to-noise ratios. They reported that the middle trials, in which distortion was neither too great nor too mild, were most sensitive to the perceptual habits of their subjects.

In another study, Rosenzweig and Postman (*15*) described the mechanism of the effect of familiarity on intelligibility somewhat as follows: when the stimulus word is masked in noise, only a part of it is discriminated, and subject's language habits tend to favor a small number of competing responses relatively high in frequency of occurrence. If the stimulus is a familiar word, it is likely to be prominent among these competing responses and is quite likely to be chosen; on the other hand, if the stimulus is of low familiarity, it is unlikely to be among the competing responses.

Viewing interaction effects in another way, emphasis might be placed upon the listener's use of available sound clues. In the frequency range where interaction effects seem most pronounced, 780 to 2 040 cps (again in terms of the filter cut-off points selected for this study), a large number of sounds are distorted to varying degrees. If a stimulus word is relatively familiar, a listener will need fewer, less distinct phonetic clues to identify it, and the clues available may be just sufficient for him to make a correct response. Conversely, if the stimulus is unfamiliar, the same clues may not suffice to permit correct identification. When the distortion becomes severe, as in the 540 cps low-pass condition, the phonetic clues are reduced to the point where even the familiar words are difficult to identify, and changes in intelligibility with familiarity become relatively less apparent.

The findings on interaction appear pertinent to high frequency hearing loss. Most clinical audiologists would agree that losses involving the middle range of frequencies are the ones for which it is most difficult to predict the discrimination score. It is not uncommon for two people with essentially

the same pattern of pure-tone loss to vary widely in discrimination ability. Perhaps it is in the middle frequency range (such as the 780 to 2 040 cps range of this study) that differences such as those in academic background, intelligence, motivation, and verbal aptitude come to the fore. Other things being equal, it would seem that a person of high intelligence and superior verbal ability would find more of the test words familiar to an extent enabling him to take greater advantage of the available phonetic clues and consequently score higher in discrimination than would a person with a lower level of intelligence and verbal ability for whom the same clues conceivably could be lost. In terms of the interaction effects of this study, as a hearing loss involves frequencies down to 540 cps and below, acoustic deprivation begins to override the advantages that can be taken of word familiarity; and to that extent the loss becomes a disproportionately greater handicap.

Correlations. Error scores on words for the 960 cps low-pass condition were compared with the logarithms of their frequencies of usage, the purpose being to investigate whether, among words already high in frequency of usage, further increases continued to be accompanied by better intelligibility. It will be recalled that list 3B (extremely familiar, contextual) was composed of contextual words ranging from a usage frequency of 708 to 52 101. The rank difference correlation of error scores for these 17 words and their logarithmic frequencies of usage was —.35, indicating a slight tendency for error scores to decrease as frequency of usage increased. For a second comparison, the 20 words of list 2B (extremely familiar) were combined with the five words in list 1C (highly familiar) having the highest frequency of usage, the entire list of 25 words covering a range of frequency usage from 1 850 to 9 499, with one word having an extreme count of 33 404. The rank difference correlation between the errors on these words and their logarithmic frequencies of usage was .16. Distributions of errors for each of the above comparisons appeared normal.

Although the small Ns prevent any sweeping statements, the essential lack of correlation between intelligibility and further increases in usage frequencies of the words already high in familiarity indicates the possibility of an asymptote in the intelligibility-familiarity relationship of words after a certain amount of familiarity is achieved. This possibility is supported by the relatively close correspondence between the means of lists 1C (highly familiar) and 2B (extremely familiar), even though their average frequency-of-usage counts differ widely. In agreement with this evidence, Howes (*10*) found no further decrease in critical speech-to-noise ratio relative to word intelligibility above usage frequencies of 2 000. On the other hand, Rosenzweig and Postman (*15*) studied words over a usage-frequency range of one to 3 000 and did not report a levelling off in intelligibility, frequency-of-usage relationships. In any event, it probably would be well to omit words of usage frequencies over 2 000 from intelligibility test lists in the interest of developing precise, predictable tools in which word familiarity can be rigidly controlled.

Summary

In studying the relationship between word familiarity and intelligibility, seven monosyllabic word lists were divided into groups of three, two, and two, the lists of each group being matched phonetically and varied systematically with respect to familiarity, defined by frequency of usage according to the Lorge count. After being recorded under seven conditions of distortion achieved by low-pass filtering, the seven lists were presented auditorily to listener groups. The 210 college students serving as listeners were divided into seven groups of 30, each group listening to all seven lists under a single condition of distortion. Test words were presented only once, with no previous practice permitted.

Lists characterized by greater familiarity, even to a slight degree, were significantly more intelligible. Findings were related to PB-50 and W-22 intelligibility lists, implications for refinement of test lists were considered, and results were tenuously generalized to persons with impaired hearing.

References

1. Black, J. W., Accompaniments of word intelligibility. *J. Speech Hearing Dis.*, 17, 1952, 409-418.
2. Dewey, G., *Relative Frequency of English Speech Sounds.* (rev. ed.) Cambridge: Harvard Univ. Press, 1950.
3. Edwards, A. L., *Experimental Design in Psychological Research.* New York: Rinehart, 1950.
4. Egan, J. P., Articulation testing methods. *Laryngoscope,* 58, 1948, 955-991.
5. Eldert, E., and Davis, H., The articulation function of patients with conductive deafness. *Laryngoscope,* 61, 1951, 891-909.
6. Fairbanks, G., *Voice and Articulation Drillbook.* New York: Harper, 1940.
7. French, N. R., and Steinberg, J. C., Factors governing the intelligibility of speech sounds. *J. acoust. Soc. Amer.,* 19, 1947, 90-119.
8. Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, Elizabeth G., Eldert, Elizabeth, and Benson, R. W., Development of materials for speech audiometry. *J. Speech Hearing Dis.,* 17, 1952, 321-337.
9. Hirsh, I. J., Reynolds, Elizabeth G., and Joseph, M., Intelligibility of different speech materials. *J. acoust. Soc. Amer.,* 26, 1954, 530-538.
10. Howes, D., On the relation between the intelligibility and frequency of occurrence of English words. *J. acoust. Soc. Amer.,* 29, 1957, 296-305.
11. Howes, D. H., and Solomon, R. L., Visual duration threshold as a function of word-probability. *J. exp. Psychol.,* 41, 1951, 401-410.
12. King-Ellison, Patricia, and Jenkins, J. J., The durational threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.,* 67, 1954, 700-703.
13. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin, 1953.
14. Postman, L., and Rosenzweig, M. R., Practice and transfer in the visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.,* 69, 1956, 209-226.
15. Rosenzweig, M. R., and Postman, L., Intelligibility as a function of frequency of usage. *J. exp. Psychol.,* 54, 1957, 412-422.
16. Silverman, S. R., and Hirsh, I. J., Problems related to the use of speech in clinical audiometry. *Ann. Oto. Rhino. Laryng.,* 64, 1955, 1234-1244.
17. Solomon, R. L., and Postman, L., Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.,* 43, 1952, 195-201.
18. Thorndike, E. L., and Lorge, I., *The Teacher's Word Book of 30,000 Words.* (2nd ptg.) New York: Teach. Coll., Columbia Univ., 1952.

Intrinsic Cues and Consonant Perception

WILLIAM S-Y. WANG

CHARLES J. FILLMORE

It has long been observed that neighboring phones influence each other in systematic ways, and that the influence which phone A has on phone B is frequently a useful secondary cue for the perception of phone A. As examples of such secondary cues in English, the shortening of vowels before unvoiced consonants and the slight nasalizing of vowels before nasals might be considered. Since such influences are predictable within English, the phonetic variation in the vowel is usually considered allophonic. For instance, the vowels in [bĭt], [bid], and [bĭn] are called allophones of the phoneme /i/. However, in most phonetic discussion, it is useful to distinguish those secondary cues which reflect the speech habits of a particular community from those which reflect the structure of the speech mechanism in general. The former is called *extrinsic* and the latter, *intrinsic*.

In the examples just given, vowel shortening would be an extrinsic type of secondary cue since this is probably a property of English, not of speech in general. Aside from the redundancy that extrinsic cues provide to aid perception, it is well known that they play a central role in the evolution of sound systems (*3*). In some American dialects, vowel duration has already become the primary cue for word-pairs such as 'ladder-latter,' 'bidder-bitter,' and similar pairs, when the /d/ and /t/ are both flaps, and such as 'his-hiss,' 'eyes-ice,' and similar pairs, when the /z/ is unvoiced.

On the other hand, the vowel nasalization is explainable in terms of the anticipatory movements of the velum without consideration of the particular language involved. The nasalization of vowels before nasal consonants would be an intrinsic type of secondary cue because it reflects a property of the speech mechanism in general. Other intrinsic cues may be exemplified by the lip-rounding of labial consonants before rounded vowels, and by the higher fundamental frequency of the high vowels due to the effect that the raising of the tongue-body has on the geometry of the larynx. Since intrinsic cues obey more or less the same constraints in all normal speech mechanisms, it is not likely that they can assume independent signal value in any language.

William S-Y. Wang (Ph.D., University of Michigan, 1960) is Research Associate, and Charles J. Fillmore (M.A., University of Michigan, 1957) is Research Assistant, Communication Sciences Laboratory, University of Michigan. This study was supported by the office of Scientific Research of the U. S. Air Force under contract AF-49-(638)-492.

TABLE 1. Confusion matrix for consonants preceding [ɛ]. Each entry in the table indicates the number of responses (10 listeners) for the corresponding intended consonant and perceived consonant as given in the left-hand margin and in the column headings, respectively.

Intended Consonant	*Perceived Consonant*								
	p	t	k	b	d	g	m	n	ŋ
p	37	32	12	1	2	1			
t	25	45	14	2		1			
k	18	29	32	2	2	2			
b	8	23	6	25	17	2			1
d	4	16	7	11	40	5			1
g	2	10	20	2	7	41			1
m							66	13	8
n							48	30	10
ŋ							39	17	33

Statement of Problem

The present study investigates the effect of some of the intrinsic secondary cues. Specifically, the purpose is to evaluate the influence that the consonant-vowel intereffects have on the perception of the consonant. From analytic considerations and from measurements it is known, for instance, that (a) high vowels have smaller amplitudes (*4*); (b) labial consonant next to [i] and alveolar consonant next to [u] usually cause the greatest bend in the second formant (*2*, *8*); and (c) vowels adjacent to nasal consonants are slightly nasalized.[1] A question may thus be asked concerning the influence these cues in the vowel might have on the perception of the consonant.

Since intrinsic cues are in a sense completely predictable regardless of the language involved, their influence in perception may not be as readily demonstrable as the influence of extrinsic cues, especially when the speech stimuli are presented under favorable circumstances. In the present experiment, masking noise was used to induce greater differences in the articulation scores for the various consonants. The purpose of the experiment was to investigate whether the above-mentioned intrinsic cues influence consonant judgments in a systematic way.

Design of Experiment

Test items were constructed of the form CVC, where C is one of the consonants [p, t, k, b, d, g, m, n, ŋ], and V is one of the vowels [i, ɛ, a, ɔ, u], with one example of every combination. The complete list of 405 (9 x 5 x 9) items was randomized and recorded on magnetic tape at roughly three-second intervals, with longer pauses separating groups of five items. In the initial position, [p, t, k] were always aspirated and [b, d, g] were not. Since previous experiments have shown that voicing is not a crucial feature in distinguish-

[1] W. D. Chapman, Nasalization of American English vowels, paper presented at 1957 convention of the American Speech and Hearing Association, Cincinnati.

TABLE 2. Confusion matrix for consonants preceding [u]. Each entry in the table indicates the number of responses (10 listeners) for the corresponding intended consonant and perceived consonant as given in the left-hand margin and in the column headings, respectively.

Intended Consonant	*Perceived Consonant*								
	p	t	k	b	d	g	m	n	ŋ
p	15	49	19						
t	3	65	15		1				
k	15	45	24		1				
b	10	41	24	1	5	1			
d	1	53	15	2	11				
g	13	44	12	2	9	2	1		
m	2	4		4	4	2	25	12	24
n				2	10		19	28	21
ŋ		1			1		23	27	28

ing [p, t, k] from [b, d, g] in initial position (5), no attempt was made to ensure that [b, d, g] were voiced initially. Final stops were unreleased.

A signal-to-noise ratio of approximately 6 db was obtained by means of a mixer. The frequencies below 200 cps and above 6 500 cps in both the Gaussian noise and the test tape were attenuated by prefiltering.

The listening task was performed by 10 phonetically trained subjects. They were informed of the set of consonants and vowels from which the test items were constructed, and were asked to transcribe the tape.

Results of Listening Tape

Responses on final consonants were nearly random. Data on these have not been included in this report.

The correct identification of vowels, with 810 correct responses possible for each vowel, was high: [ɛ], 95.1%; [u], 90.9%; [ɔ], 88.6%; [a], 88.4%; and [i], 83.9%.

TABLE 3. Confusion matrix for consonants preceding [ɔ]. Each entry in the table indicates the number of responses (10 listeners) for the corresponding intended consonant and perceived consonant as given in the left-hand margin and in the column headings, respectively.

Intended Consonant	*Perceived Consonant*								
	p	t	k	b	d	g	m	n	ŋ
p	29	17	21	2	1	3	1		
t	23	23	28	2	1				
k	25	20	22	5	1		1		
b	20	23	12	13	6	4			
d	1	37	6	2	30	6			
g	14	30	18	12	5	3		1	
m				1			64	7	14
n	1				1		10	43	29
ŋ							19	16	45

TABLE 4. Confusion matrix for consonants preceding [a]. Each entry in the table indicates the number of responses (10 listeners) for the corresponding intended consonant and perceived consonant as given in the left-hand margin and in the column headings, respectively.

Intended Consonant	*Perceived Consonant*								
	p	t	k	b	d	g	m	n	ŋ
p	9	36	20	3	1				
t	4	52	22	1					
k	1	43	25	3		1			
b	12	12	6	39	7	5			
d		21		4	48	9			
g		18	3	3	35	22		1	
m							74	2	9
n							2	57	22
ŋ			1				2	37	44

The data on initial consonants, presented in Tables 1 through 5 by means of 9 x 9 matrices, are taken only from responses in which the vowels have been correctly identified. The sum of the entries in these tables represents a total of 3 619 responses, corresponding to the total of correct vowel responses. The total varies from one row to another because consonants preceding incorrectly perceived vowels were not counted. The intended consonant is indicated in the left margin and the consonant response is indicated in the column headings. The number of correct identifications is entered along the main diagonal from upper left to lower right.

Interpretation of Results

The articulation scores for all consonants pooled are considerably lower for the high vowels than for the low vowels: [a], 45.7%; [ɛ], 43.1%; [ɔ], 33.6%; [i], 25.3%; and [u], 24.6%.

TABLE 5. Confusion matrix for consonants preceding [i]. Each entry in the table indicates the number of responses (10 listeners) for the corresponding intended consonant and perceived consonant as given in the left-hand margin and in the column headings, respectively.

Intended Consonant	*Perceived Consonant*								
	p	t	k	b	d	g	m	n	ŋ
p	34	20	12	3			1		
t	34	20	18	3					
k	24	30	12	11	1				
b	33	10	4	21	4				
d	16	18	9	16	12	4			
g	20	14	16	12	1	1			
m	1	1		4			60	10	9
n				3	3		38	22	20
ŋ	1						41	9	23

TABLE 6. Articulation scores in per cent for consonants pooled by place of production.

Place	a	ɛ	ɔ	i	u
Bilabial	51.91	50.39	44.73	50.66	16.94
Alveolar	64.88	44.40	39.51	22.88	42.78
Velar	38.24	41.25	29.54	16.11	21.77

This seems to suggest the thesis that consonant-vowel intereffects provide a stronger secondary cue when the intrinsic amplitude of the vowel is greater. Since many consonants depend heavily on formant transitions for their identification, greater vowel amplitude implies a better opportunity to perceive the transitions and to identify the preceding consonant. This effect is all the more important in the present experiment because it is reasonable to believe that the Gaussian noise masks the noise-like aspirations more effectively than it does the transitions. Therefore, the value of the cues in the aspiration for both manner and place of articulation is more greatly reduced than that of the formant cues. To a reasonable extent, this type of intrinsic cue would probably increase in importance with an increase in noise level.

Table 6 shows the articulation scores for each vowel with the consonants pooled as to place of articulation. Thus the entry of 51.9% (derived from 122/235 in the [a] column and the bilabial row indicates the proportion of correctly identified [p, b, m] preceding correctly identified [a]. The articulation scores for bilabials before [i] and for alveolars before [u] are impressively higher than the other scores for the same vowels, respectively (Table 6). It should be noted here that the point of interest is in the relative differences among scores within individual vowel columns; the absolute values of the scores again, as for all consonants pooled, reflect the dependence of the articulation scores on vowel amplitude. It can be seen that in the [i] column, the bilabial articulation score is greater than the sum of the other two scores; and the same is true for the aveolar score in the [u] column. Such a condition does not obtain for the three more central vowels. These results may be explained by the earlier mentioned consideration that labials before [i] and alveolars before [u] give the greatest bends in the second formant. Other things being equal, the degree of bend in F_2 is positively correlated with the identifiability of its associated consonant.

Table 7 shows the articulation scores for each vowel with consonants pooled

TABLE 7. Articulation scores in per cent for consonants pooled by manner of production.

Manner	a	ɛ	ɔ	i	u
Aspirated Plosive	39.91	44.36	32.89	29.60	41.27
Unaspirated Plosive	44.49	42.57	18.93	16.67	5.67
Nasal	70.00	48.90	60.80	42.86	34.18

by manner of articulation. Except before [u], the nasal consonants have the highest scores; and except before [a], the aspirated plosives have the second highest scores. The general ranking seems to agree with power measurements made from conversational speech (*6*), that is, aspirated plosives have greater power than unaspirated plosives but less than nasal consonants. The high score for the nasals may be additionally accounted for by the secondary cue of nasalization in the adjacent vowel (see footnote 1).

As an inspection of Tables 1 through 5 will show, the errors in manner of articulation are mostly in identifying unaspirated plosives as aspirated. There are very few confusions between the latter and the nasals. The identification of nasals as a class is nearly complete.

Discussion

There are several issues exhibited in the results which need further investigation. While the consonant articulation scores are high for those items with great formant bands, it was observed that there were also many wrong responses in the same direction; that is, before [i] and [u] many consonants were wrongly identified as bilabial and alveolar, respectively. The agreement may be regarded as a contradiction from one point of view since the more certain a subject is about the identity of a class of objects, the less likely should he be to misidentify objects as belonging to that class. On the other hand, there may be something about the experimental situation that induces the subject to repeat the symbols that he has already used.

In line with the latter point of view, it was observed that before [a], for instance, among items heard as aspirated plosives, three subjects showed a high preference for velars while seven subjects chose predominantly alveolars. The observation that alveolars were given a high response score may be an accident of the selection of subjects where each subject has his preferred symbol for items about which he is completely uncertain. It is likely, however, that such issues, including the difference in ranking of the vowels by vowel articulation scores or by articulation scores of adjacent consonants, would be resolved if the experiment were to be expanded in scope and an additional symbol were used for items about which the subjects are completely uncertain. In planning such an expanded experiment, much of the time and labor involved in the evaluation of the responses and the construction of the confusion matrices could be done with aid of a digital computer.

The results of this pilot experiment seem to demonstrate the dependence of consonant perception on the intrinsic cues present in the vowels, although the nature of this dependence has been controversial (*1*, 7, *9*). In the ranking of vowels by articulation scores of preceding consonants, for instance, there is only one chance in 10 for the two high vowels to take the last two ranks, but the resulting ranking may be expected from previous measurements of vowel amplitudes. Similarly, the degree of formant bend and the nasalization in the vowel seem to have influenced the corresponding scores of the identification of the preceding consonants.

Summary

A distinction has been made between secondary cues which are extrinsic and those which are intrinsic. The purpose was to investigate the contribution of some intrinsic cues in consonant perception. Subjects were 10 phonetically-trained listeners. Test items were 405 consonant-vowel-consonant syllables consisting of nine consonants and five vowels in all combinations. Results of correct identification of initial consonants suggest that vowel amplitude, degree of formant bend, and vowel nasalization are significant parameters in the vowel for identifying the consonant which precedes it.

Acknowledgment

The authors wish to thank Dr. Peter Ladefoged of the University of Edinburgh for serving as the speaker in preparing the test tape.

References

1. Delattre, P., Unreleased velar plosives after back-rounded vowels. *J. acoust. Soc. Amer.*, 30, 1958, 581-582.
2. Delattre, P. C., Liberman, A. M., and Cooper, F. S., Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, 27, 1955, 769-773.
3. Kurath, H., The loss of long consonants and the rise of voiced fricatives in Middle English. *Language*, 32, 1956, 435-445.
4. Lehiste, Ilse, and Peterson, G. E., Vowel amplitude and phonemic stress in American English. *J. acoust. Soc. Amer.*, 31, 1959, 428-435.
5. Reeds, J., and Wang, W. S-Y., The perception of stops after [s]. *Phonetica*, 6, 1961.
6. Sacia, C. F., and Beck, C. J., The power of fundamental speech sounds. *Bell Syst. tech. J.*, 5, 1926, 393-403.
7. Sherman, Dorothy, The influence of vowels on recognition of adjacent consonants. *J. Speech Hearing Dis.*, 17, 1952, 198-212.
8. Stevens, K. N., and House, A. S., Studies of formant transitions using a vocal tract analog. *J. acoust. Soc. Amer.*, 28, 1956, 578-585.
9. Wang, W. S-Y., Transition and release as perceptual cues for final plosives. *J. Speech Hearing Res.*, 2, 1959, 66-73.

Binaural Hearing Aids and Speech Intelligibility

JAMES JERGER

RAYMOND CARHART

DONALD DIRKS

The recent advent of wearable binaural hearing aids has already generated a conflicting literature on the extent of additional benefit that the hearing impaired individual may expect to derive from binaural versus monaural amplification.

Several authors (*1, 2, 3, 5,* 7, *10*) have cited both theoretical arguments and clinical observations supporting the thesis that either some or all hearing-aid users may expect anywhere from mild to substantial additional benefit from binaural as opposed to monaural hearing aid use.

On the other hand attempts to demonstrate this advantage objectively (*4, 6, 11*) have, with one exception (*1*), met with little material success.

The present status of the problem has been eloquently summarized by Haskins and Hardy (*5*):

> By and large, routine clinical tests are not effective measures of the improvement of function; . . . . All of /the/ cases studied with binaural hearing profited sufficiently to choose it. They enjoyed the quality, and, like others who enjoy the effect, they report that events in sound are more natural, not so 'thin,' with 'both sides of the head alive.' In general they localize better and hear with more assurance, and with obvious relief from communicative pressures. In most instances they perform better in the presence of background noise than they have ever been able to do with monaural hearing.

Haskins and Hardy go on to emphasize the subjective nature of this improved performance:

> . . . in some instances the test-variables, with and without binaural aids, were so slight that no quantifiable significance could be derived; in other instances, the variations between monaural and binaural hearing were astonishingly great. In either extreme, there were no clear indicators of cause and effect. It is clear that localization, *per se,* is by no means the factor of greatest importance; it is rather a matter of more information received and a sensation of awareness and security in listening which comes from the ready perception of depth-in-sound, which these persons report. They are more relaxed, more assured, and *more effective communicators* (italics ours).

It is this last description, 'more effec-

James Jerger (Ph.D., Northwestern University, 1954) is Associate Professor of Audiology and Director of Audiological Research, Northwestern University. Raymond Carhart (Ph.D., Northwestern University, 1936) is Professor of Audiology and Director of the Program in Audiology, Northwestern University. Donald Dirks (M.S., University of Kansas, 1959) is a Research Assistant in Audiology, Northwestern University. This research was supported by research grant VM1005M-1926 from the Prosthetic and Sensory Aids Service of the Veterans Administration.

tive communicators,' with which the present series of studies is directly concerned. Presuming that the type of hearing aid used probably exerts no significant influence on the user's expressive communicative capacity, then the question arises as to whether he is a more effective communicator because he understands other people's speech better. Can it be demonstrated that binaural hearing aids actually improve speech intelligibility, or does the basis for 'more effective communication' lie in some other, less easily defined and measured, set of variables?

The sincerity of the enthusiastic reports on the efficacy of binaural hearing aids from hearing-aid users, dealers, and clinical audiologists, cannot be doubted. But the cautious scientist, ever mindful of the vagaries of subjective reports, and of how often they have led to false conclusions in other areas, inevitably seeks quantifiable measures to substantiate subjective reports. In their absence he has no choice but to maintain a guarded attitude until and unless such measures can be devised.

The purpose of the present study was to seek objective verification of subjective reports that binaural hearing aids improve the user's ability to understand speech in difficult listening situations; specifically in the presence of competing speech or noise, and in noisy environments when the source of speech is not stable. To this end, two quantifiable test procedures were devised. Both tests involved a primary signal from one direction, to which the listener must attend in the presence of a secondary or competing signal from another direction. In one procedure, Northwestern University Test #2, the primary signal was a PB-50 word and the secondary signal was a complete sentence. In the second procedure, NU Test #3, the primary signal was a complete sentence and the secondary signal was continuous discourse. Both procedures were specifically designed to re-create, as closely as possible within the confines of quantifiable measurement, the kinds of listening situations in which the reputed advantage of the binaural hearing aid ought to be maximal, according to subjective reports available to the writers.

Test Materials

NU Test #2 consists of 15 non-duplicating forms. Each version is a two-channel tape recording in which 50 PB words are recorded on one channel and 50 sentences from the Bell Telephone Intelligibility Lists are recorded on the other channel. Items are recorded in such sequence that each PB word begins slightly after the initiation of its corresponding competing sentence. PB words recur at 10-sec intervals, and all are preceded by the carrier phrase, 'Write the word. . . .'

The PB words were recorded by a male talker with General American dialect, the competing sentences by a different male talker, also with General American dialect.

In the present experiment only three of the 15 forms of Test #2 were actually employed. They were selected on the basis of their demonstrated equivalence on both normal and impaired ears during previous research.

NU Test #3 is a modified sentence intelligibility test with three equivalent forms. Each form consists of a two-

channel tape recording in which 30 questions and commands are recorded on one channel, and the simultaneous discourse or running speech of two separate talkers is recorded on the other channel. The 90 sentence items required for the three versions of the test were drawn from a sentence intelligibility test originally developed at the Psychoacoustic Laboratory of Harvard University by Karlin and others (*9*), and known, in the literature, as PAL Auditory Test #8. This is essentially a multiple choice intelligibility test in which the subject hears a simple sentence either asking a question or giving a command, and then must select the correct answer from among four possible choices.

Tape recordings for NU Test #3 were made in such a way that approximately three seconds after each sentence item the next item might or might not be switched over to the opposite channel. Since successive items recurred at 4-sec intervals, switching occurred approximately one second before the next item began. The switching sequence over each block of 30 items was determined by a table of random numbers with the single constraint that the test items must appear an equal number of times (15) on each of the two channels. A separate and distinct random sequence was employed for each of the three test forms.

The purpose of this random switching sequence was to create a relatively unstable listening situation. When the tape is played through a two-channel amplifying system such that each channel feeds a separate loudspeaker, the subject hears the sentence (primary signal) from one loudspeaker and the continuous discourse (competing signal) from the other loudspeaker. Over the 30 item sequence, however, the position of the primary signal shifts, first to one loudspeaker, then to the other, in relatively unpredictable fashion. As a consequence the spatial location of the primary signal with respect to the listener varies from item to item in a nonsystematic manner.

The secondary or competing message of Test #3 consisted of the simultaneous running speech of two voices, one male, the other female. In making the original recording the male voice read from a basic textbook on hearing, the female voice from a textbook on the history of philosophy. Both talkers were instructed to read at an average, comfortable rate and to keep their voices peaked to 0 on a VU meter as they read.

Apparatus. All tapes, both for Test #2 and Test #3, were recorded and played back on a two-channel tape recorder (Berlant, Model 30). The dual outputs from the tape recorder were sent through a console, with provision for independent amplification and attenuation of each channel, then fed to two matched loudspeakers

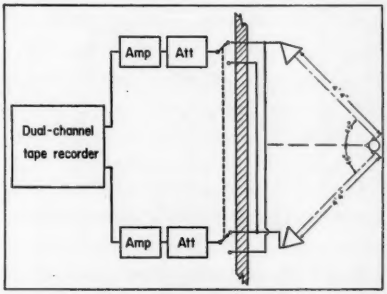

FIGURE 1. Simplified block diagram of experimental apparatus.

in the test chamber. Figure 1 shows the arrangement of the basic apparatus. All equipment, except for the two loudspeakers and a set of earphones, was located in a control room adjacent to but well isolated from the test chamber. The subject was seated in the test chamber at a point equidistant from the two loudspeakers. Each loudspeaker faced the subject in a plane displaced 45° from his median plane.

For both tests the primary to secondary ratio at the two loudspeakers was always 0 db. That is, primary and secondary signals were always of equal intensity. The intensity level of all speech signals was held constant at 65 db SPL (equivalent SPL of thermal noise presented through a single speaker and measured on the C scale of a sound level meter at the position of the subject's head in the sound field) for both tests and under all experimental conditions.

Hearing Aids. For each subject speech intelligibility was measured under three conditions of amplification: binaural (A), monaural-head (B), and monaural-body (C). To achieve these separate conditions three matched hearing aids were employed. All were Beltone instruments (Fortissimo model) with T-10 receivers. Frequency response curves of all three aids were essentially equivalent. They are illustrated in Figure 2. Two of the aids were permanently mounted in a headband which, when positioned on the subject's head, located each aid just above the ear. This arrangement was used for both binaural (A) and monaural-head (B) conditions. For the binaural condition both aids were turned on. For the monaural-head condition one aid was turned off and its earpiece removed from the subject's external auditory canal. The third aid was placed in a cloth halter which fitted comfortably around the subject's head and positioned the aid approximately at the center of the chest. This arrangement was used for the monaural-body (C) condition. In all three aids the tone control adjustment was kept constant in the 'flat' position.

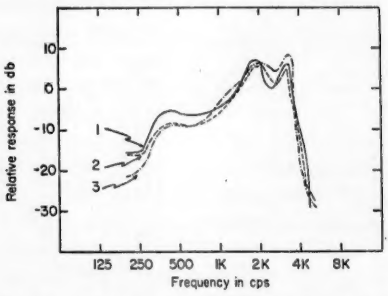

FIGURE 2. Frequency response curves of the three hearing aids. Aids 1 and 2 were mounted in the binaural headset. Aid 3 was mounted on the center of the chest.

Subjects. A total of 48 subjects with bilateral sensorineural hearing loss were tested in this experiment. They were divided into two groups of 24 subjects each on the basis of chronological age. Group I included only subjects under the age of 51 years, and Group II included only subjects over the age of 59 years.

Of the 24 subjects in Group I, 11 were male and 13 were female. The median age for this group was 42 years with a range from 19 to 50 years.

Group II was composed of 10 males and 14 females. The median age for this group was 67 years with a range from 60 to 76 years.

Subjects were selected from the files of patients who had received hearing aid evaluations at one of the hearing clinics of Northwestern University. In order to ensure a relatively homogeneous population of potentially successful binaural hearing aid users, the following selection criteria were employed. First, the hearing loss must be of the sensorineural type (interweaving air and bone conduction thresholds) and relatively bilaterally symmetrical. Specifically a maximum difference of only 12 db was permitted between spondee thresholds on the two ears. Furthermore, both spondee threshold losses must fall within the range from 25 to 62 db.

As a result of this selection procedure the typical subject showed either a flat or gradually falling audiometric contour characterized by close bilateral symmetry and relatively good to excellent PB discrimination on either ear.

Of the 48 subjects, 32 were monaural hearing aid users who had worn aids for periods ranging from five months to eight years. Two others had worn binaural hearing aids for a period of about four years, and the remaining 14 used no aid at all.

Procedure

Each subject was tested in a single experimental session lasting about two and one-half hours. Routine pure-tone and speech audiometry preceded actual tests of speech intelligibility under the various conditions of amplification.

These routine tests included a standard pure-tone air conduction audiogram, a test for sensorineural acuity level (SAL), monaural spondee thresholds, and the administration of a PB-50 word list separately to each ear at a level 40 db above the spondee threshold. Both spondee and PB materials for this routine speech audiometry were prerecorded on magnetic tape and played back through the same tape recorder and console used to deliver NU Tests #2 and #3 to the loudspeakers. For routine speech audiometry, however, all materials were administered by earphone (Telephonic, type TDH-39).

Following these standard audiometric measures, the earphones were removed from the subject's head, and the appropriate hearing aid condition was set up. The order in which successive conditions (binaural, monaural-head, and monaural-body) were tested was counterbalanced over the two experimental groups in such a way that each condition occurred an equal number of times in each possible sequential position. Within a particular condition NU Test #2 occurred first exactly half of the time and NU Test #3 occurred first exactly half of the time. Finally, the employment of the three alternate forms of each test was counterbalanced in such a way that each form occurred an equal number of times in each of the three possible positions of order.

In selecting the ear to be fitted under the two monaural conditions the experimenter exercised his clinical judgement as to which ear was the more appropriate to fit, on the basis of audiometric contour and unaided PB discrimination. In most cases the choice between ears was quite arbitrary since losses were well balanced and PB discrimination was relatively good on either ear.

After the appropriate hearing aid condition had been set up and the aid or aids turned on, the subject was instructed as follows:

> Now you are going to hear some voices coming through the loudspeakers. I want you to adjust the little wheel on the hearing aid until the voices are at a level where you would like to listen to them; not too loud, but not too faint either. Try to make it just comfortable. When you have it set where you like it, just tell me.

In the binaural condition these additional instructions were given:

> This time there are two hearing aids to adjust. Adjust this one (right ear) first, until it is comfortable. Then leave it set where it is, and adjust this other one (left ear) until it is comfortable. Try to make the voices sound the same in each ear so that it is not louder on one side than on the other.

One minute of prerecorded continuous discourse (taken from the actual secondary or competing signal of Test #3) was then played at an SPL of 65 db.

This gain setting procedure was followed just prior to the administration of Test #2 and Test #3 under each condition of amplification. As soon as the subject reported a satisfactory adjustment, first one then the other test was given.

In the case of NU Test #2 the primary signal (PB word) always came from the speaker opposite the ear fitted with the monaural aid. In the case of NU Test #3 this distinction was irrelevant since exactly half of the primary items came from one loudspeaker and half from the other.

NU Test #2 was administered by simply playing the appropriate tapes, at an SPL of 65 db, and having the subject repeat back each PB word over a talkback system. Responses were scored as either correct or incorrect by the experimenter at that time. NU Test #3 was administered in paper and pencil fashion by having the subject mark the correct answer to each multiple choice item on a printed answer sheet.

In all three hearing aid conditions the subject was free to move his head as much as he desired. Rest periods were interspersed between successive experimental conditions to avoid, as much as possible, subject fatigue during these relatively time consuming and tedious tests.

Results

Basic Audiometric Data. Table 1 shows the distribution of basic audio-

TABLE 1. Distribution of average pure-tone hearing level (500 to 2 000 cps) in db, spondee threshold hearing level in db, and PB discrimination score in percent for all subjects.

Measure	*Ear*	*Group I (Young)*		*Group II (Old)*	
		Median	*Range*	*Median*	*Range*
Average Pure-Tone Level	R	42	30-62	45	25-53
	L	40	28-58	39	30-55
Spondee Level	R	47	34-62	45	28-58
	L	43	34-64	42	32-60
PB Score	R	86	66-94	84	44-96
	L	86	60-96	84	62-96

metric results in both experimental groups. It is readily apparent that, as a result of the selection procedures described above, the typical subject in this experiment was almost hand picked as an 'ideal' candidate for binaural amplification. He had a mild to moderate bilaterally symmetrical loss with relatively good PB discrimination. The characteristics of his hearing loss were, as a matter of fact, purposely chosen to be 'optimal' for binaural hearing aid use (a flat or gradually sloping moderate bilateral loss with good speech discrimination ability). This procedure was followed in order to ensure that any inherent advantage of binaural over monaural amplification would not be obscured by contamination of the experimental groups with subjects whose hearing loss presented special problems inimical to the successful use of binaural hearing aids.

The distributions in Table 1, then, are not typical of results to be expected in a truly random sample of subjects with sensorineural hearing loss. They reflect, instead, a purposeful selection procedure designed to assure balanced residual hearing and relatively good speech discrimination ability.

TABLE 2. Median and range of percent correct scores on NU Test #2 for three conditions: monaural-body (C), monaural-head (B), and binaural (A).

Group	*Measure*	*Condition*		
		Mon-Body (C)	*Mon-Head (B)*	*Bin (A)*
I	Median	12	13	27
	Range	0 to 42	0 to 52	2 to 54
II	Median	21	22	27
	Range	0 to 62	0 to 62	4 to 54

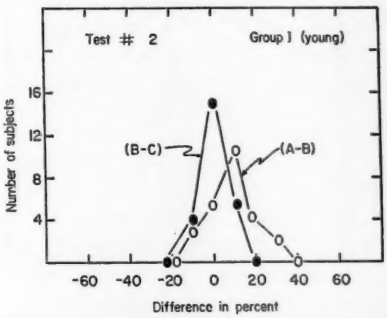

FIGURE 3. Distribution of A—B and B—C for NU Test #2 in Group I (young) subjects.

NU Test #2. In Test #2 (words and competing sentences), the primary signal was a PB word from one loudspeaker, and the secondary, or competing signal, was a complete sentence simultaneously emitted from the other loudspeaker.

Table 2 shows the medians and ranges of scores obtained on this test in both experimental groups.

In order to compare monaural and binaural performance most meaningfully, two difference scores were computed for each subject. First, the score obtained under the monaural-body condition (C) was subtracted from the score obtained under the monaural-head condition (B). The resultant difference score (B – C) is the extent to which performance was improved by moving a single hearing aid from the center of the chest to a position on the head above one ear. Next the score obtained under the monaural-head condition (B) was subtracted from the score under the binaural condition (A). The resultant difference score (A – B) is the extent to which performance was improved by adding a second hearing aid on the head. The

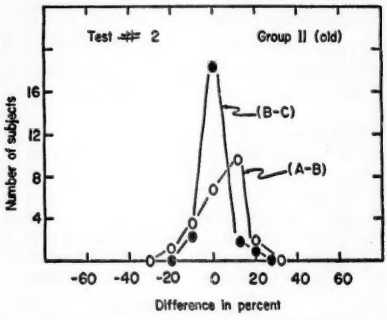

FIGURE 4. Distribution of A—B and B—C for NU Test #2 in Group II (elderly) subjects.

meaning of these difference scores may be summarized as follows for convenient reference: B — C is improvement obtained by moving single hearing aid from body to head; A — B is improvement obtained by adding second hearing aid to head.

Figure 3 shows the distribution of these two indices in experimental Group I (under 50 years). In this and all subsequent figures, distributions reflect frequency of occurrence of difference scores in class intervals of 10%. The point for 0% difference, for example, represents the frequency of occurrence of scores within the class interval from —5 to +5%. The difference labelled 20% represents the interval from 15 to 25%, etc.

Inspection of Figure 3 shows that the distribution of B — C is fairly symmetrical around 0%, indicating no apparent advantage of head versus body location for a single hearing aid, relative to performance on this specific listening task (PB words and competing sentences). The distribution of A — B is also reasonably symmetrical, but displaced slightly in the positive direction, implying a slight advantage (roughly 10%) for binaural over monaural head amplification, again on this specific listening task.

Figure 4 shows analogous distributions for Group II (over 60 years). The over-all pattern is fairly close to that of Group I, except that the A — B distribution is slightly skewed and suggests a somewhat less clear-cut advantage for the binaural system. Again, the distribution of B — C shows no advantage for a head- as opposed to a body-borne single hearing aid.

One possible reason for the lack of marked binaural superiority in the foregoing distributions is the fact that a subject with only mild hearing loss may be functioning, as a result of the auditory cues he receives on the unaided ear, as a binaural listener even though he wears only one aid. Thus binaural aids may offer him little additional advantage beyond what a single aid and an open, unaided ear already provide for him. If this were true one would expect to find, in the present data, that subjects with the greatest hearing loss reflected more binaural

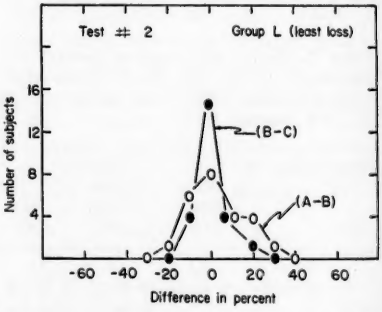

FIGURE 5. Distribution of A—B and B—C for NU Test #2 in Group L (least loss) subjects.

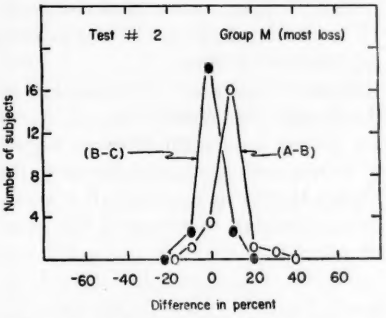

FIGURE 6. Distribution of A—B and B—C for NU Test #2 in Group M (most loss) subjects.

superiority than subjects with the least hearing loss.

In order to investigate this possibility all 48 subjects in Groups I and II were pooled and redivided into two new groups, L and M. Group L consisted of the 24 subjects with the least hearing loss for speech (average of monaural spondee thresholds). Group M consisted of the remaining 24 subjects with the greatest hearing loss for speech. Each new group was almost equally divided between former Group I and II subjects. Group L had 11 former Is and 13 former IIs. Group M had 13 former Is and 11 former IIs. Figures 5 and 6 show the result of this analysis. The predicted trend does, in fact, appear. Group L subjects (least loss) show no apparent binaural advantage while Group M subjects show a very clear positive displacement of the A – B distribution relative to the B – C distribution.

These results would appear to support the hypothesis that the subject with mild loss has less to gain from a binaural hearing aid than the subject with a moderate or severe loss, probably because he is already a binaural listener, even with just one hearing aid. Again, however, results can be interpreted only in relation to this specific listening task.

NU Test #3. In Test #3 (sentences and competing discourse) an effort was made to create a more realistic, life-like listening situation than was the case in Test #2. The primary message was a complete sentence which either asked a question or gave a command to which the subject responded by marking one

TABLE 3. Median and range of percent correct scores on NU Test #3 for three conditions: monaural-body (C), monaural-head (B), and binaural (A).

Group	*Measure*	*Condition*		
		Mon-Body (C)	*Mon-Head (B)*	*Bin (A)*
I	Median	60	60	77
	Range	0 to 93	3 to 100	27 to 97
II	Median	58	65	65
	Range	0 to 90	0 to 97	7 to 97

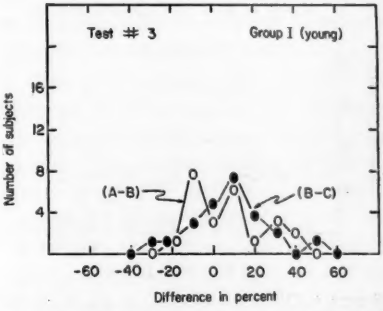

FIGURE 7. Distribution of A—B and B—C for NU Test #3 in Group I (young) subjects.

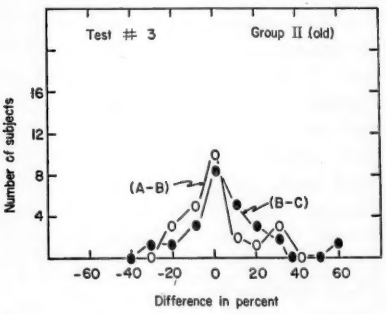

FIGURE 8. Distribution of A—B and B—C for NU Test #3 in Group II (elderly) subjects.

of four possible choices on an answer sheet. The secondary, or competing, signal was a recording of continuous discourse by two simultaneous talkers, one male and the other female. In addition, the two messages switched loudspeaker positions in a random sequence between successive test items. The purpose of this switching procedure was to simulate the real life listening situation in which the listener attempts to follow a conversation among several talkers who vary in position with respect to the listener.

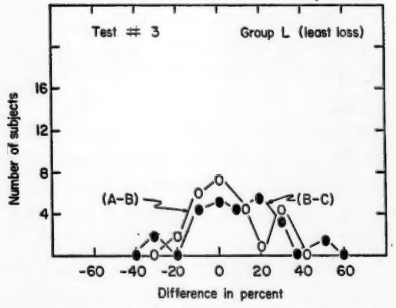

FIGURE 9. Distribution of A—B and B—C for NU Test #3 in Group L (least loss) subjects.

Table 3 shows the median and range of scores obtained on this test in both experimental groups.

Figure 7 shows the distribution of B—C and the distribution of A—B for Group I (young) listeners. Figure 8 shows analogous distributions for Group II (elderly) subjects. It is somewhat surprising to note that the overall pattern of results is quite different from that of Test #2. In Group I the B—C distribution is fairly symmetrical, but displaced about 10% in the positive direction, suggesting a slight advantage for the head as opposed to the

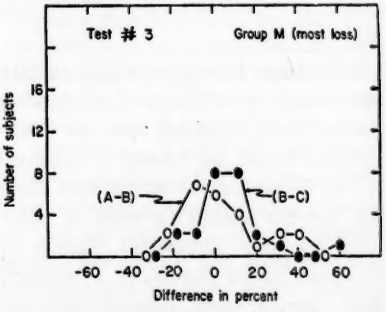

FIGURE 10. Distribution of A—B and B—C for NU Test #3 in Group M (most loss) subjects.

body location of a single aid, in this particular listening task. Of further interest is the fact that the A—B distribution is inconclusive. The mode is displaced about 10% in the negative direction but half of the cases are displaced as much or more in the positive direction. In Group II the two distributions are more orderly. Both distributions have their modes at 0% but, again, there appears to be a very slight improvement when the single aid is moved from the body to the head and

a slight deterioration in performance when the second aid is added to the head.

An alternative, and equally tenable, interpretation of these data is that they show no real differences among the three experimental conditions.

Under either interpretation there is a rather striking lack of evidence for any advantage to two hearing aids over one in this specific listening task, which involved understanding instructions against a babble of background speech.

Figures 9 and 10 show the result of regrouping all subjects on the basis of degree of hearing loss. Here there is little to distinguish Group L (least loss) from Group M (greatest loss). In both groups there is a slight improvement when the single aid is moved from the body to the head (B – C) but slightly poorer performance when the second aid is added to the head (A – B). Degree of hearing loss appears to exert no significant effect on the various distributions.

Discussion

In general, the results summarized above fail to demonstrate the marked superiority of binaural amplification which one has been led to expect from clinical reports. On Test #2 there was a slight (roughly 10%) improvement in PB discrimination against a background of competing sentences, but on Test #3 there was virtually no improvement in sentence intelligibility against a background of competing continuous discourse.

These results would be disturbing enough if they stood alone. They climax, however, a series of experiments from both this and other laboratories all pointing to the same inevitable conclusion. Binaural amplification produces little or no objectively demonstrable improvement in the ability to understand speech in quiet or in noise, against competing sentences or competing discourse, whether the speech material be isolated words or meaningful sentences. In spite of theoretical considerations which would imply significant advantages for a true binaural amplifying system, in spite of some patients' reports that they 'like' binaural aids much better than monaural aids, in spite of subjective clinical observations of dramatic improvement in the performance of some patients with binaural aids, in spite of more than six years of extensive experimentation throughout the country, there remains little concrete evidence that binaural amplification actually produces any significant improvement in the ability to understand speech.

The only reasonably verified positive finding is the observation by DiCarlo and Brown (*4*) that the ability to localize direction of a noise burst in acoustic space is improved when two hearing aids are worn. Whether this capacity has any bearing on the ability to understand speech through hearing aids under difficult listening conditions remains to be demonstrated.

The present discrepancy between results on Test #2 and on Test #3 may be related to this gross localization factor. The present feeling of the authors is that in the relatively rigidly structured acoustic space of Test #2, in which the primary signal always comes from one loudspeaker and the competing signal always comes from the other, the slight advantage in gross

localization ability afforded by the second hearing aid permitted somewhat better separation of primary and competing signals, improving everyone's score to the same minor extent, and thereby shifting the A – B distribution upward slightly in the positive direction.

In the more chaotic acoustic space of Test #3, however, where the direction of primary and competing signals was essentially unstable, the slight improvement in gross localization ability afforded by the second aid was of negligible value in the required listening task.

Of further interest in this regard are the findings of Jongkees and Veer (*8*) who found that unilateral hearing loss, per se, did not necessarily impair directional hearing, and the theoretical consideration of Wright (*12*) who questions the basic assumption that binaural hearing aids necessarily restore 'balanced' hearing to the hearing impaired.

Summary

Two indices of speech intelligibility were obtained from 48 subjects with sensorineural hearing loss under conditions of both monaural and binaural hearing-aid amplification. Results failed to reveal any appreciable advantage for two hearing aids over one. Some theoretical considerations implied by these findings are discussed.

References

1. Belzile, M., and Markle, D. M., A clinical comparison of monaural and binaural hearing aids worn by patients with conductive or perceptive deafness. *Laryngoscope*, 69, 1959, 1317-1323.
2. Bergman, M., Binaural hearing. *Arch. Otolaryng.*, 66, 1957, 572-578.
3. Carhart, R., The usefulness of the binaural hearing aid. *J. Speech Hearing Dis.*, 23, 1958, 41-51.
4. DiCarlo, L., and Brown, W., The effectiveness of binaural hearing for adults with hearing impairments. *J. auditory Res.*, 1, 1960, 35-76.
5. Haskins, Harriet L., and Hardy, W. G., Clinical studies in stereophonic hearing. *Laryngoscope*, 70, 1960, 1427-1432.
6. Hedgecock, L. D., and Sheets, B. V., A comparison of monaural and binaural hearing aids for listening to speech. *Arch. Otolaryng.*, 68, 1958, 624-629.
7. Hirsh, I. J., Binaural hearing aids: a review of some experiments. *J. Speech Hearing Dis.*, 15, 1950, 114-123.
8. Jongkees, L. B. W., and Veer, R. A. V. D., Directional hearing capacity in hearing disorders. *Acta Otolaryng.*, 48, 1957, 465-474.
9. Karlin, J. E., Abrams, M. H., Sanford, F. H., and Curtis, J. F., Auditory tests of the ability to hear speech in noise. *OSRD Report No. 3516, National Defense Research Committee*, 1944.
10. Koenig, W., Subjective effects in binaural hearing. *J. acoust. Soc. Amer.* 22, 1950, 61-62.
11. Markle, D. M., and Aber, W., A clinical evaluation of monaural and binaural hearing aids. *Arch. Otolaryng.*, 67, 1958, 606-608.
12. Wright, H. N., Binaural hearing and the hearing impaired. *Arch. Otolaryng.*, 70, 1959, 485-494.

Hearing in Children: Acoustic Environment and Audiometer Performance

ELDON L. EAGLES

LEO G. DOERFLER

Eldon L. Eagles (M.D., C.M., Dalhousie University, 1936; Dr. P.H., Johns Hopkins University, 1958) is Associate Research Professor in Maternal and Child Health, Graduate School of Public Health, University of Pittsburgh; Executive Director, Subcommittee on Hearing in Children, Committee on Conservation of Hearing, American Academy of Ophthalmology and Otolaryngology; and Director, Study of Hearing in Children. Leo G. Doerfler (Ph.D., Northwestern University, 1948) is Professor of Audiology, and Director, Department of Audiology, School of Medicine, University of Pittsburgh; and Consultant in Audiology, Study of Hearing in Children. Financial support is provided by grant number B2375 (C3) from the National Institute of Neurological Diseases and Blindness, and by funds from the American Academy of Ophthalmology and Otolaryngology to the Subcommittee on Hearing in Children. Members of the Subcommittee on Hearing in Children are Raymond E. Jordan, M.D., Chairman, Clinical Professor of Otology, University of Pittsburgh School of Medicine; John E. Bordley, M.D., Professor of Otolaryngology, Johns Hopkins University School of Medicine, and Adjunct Professor of Environmental Medicine, Johns Hopkins University School of Hygiene and Public Health; Victor Goodhill, M.D., Professor of Surgery (Otology), University of Southern California, and Consultant, John Tracy Clinic; Hollie E. McHugh, M.D., Associate Professor of Otolaryngology, McGill University; and S. Richard Silverman, Ph.D., Director, Central Institute for the Deaf, and Professor of Audiology, Washington University. Financial support in addition to that of the Subcommittee on Hearing in Children is provided by a grant from the Children's Bureau through the Commonwealth of Pennsylvania Department of Health to the University of Pittsburgh.

For several years, the American Academy of Ophthalmology and Otolaryngology, through its Subcommittee on Hearing in Children of the Committee on Conservation of Hearing, has been concerned with a long-term, nationwide study of hearing problems in children. An initial study is being conducted in Pittsburgh, through the cooperation of the Subcommittee with the Maternal and Child Health Section of the University of Pittsburgh's Graduate School of Public Health, to evaluate methods for the testing of hearing in children in order to decide which are more efficient, economical, and valid in terms of time and manpower, and to evaluate and define medical signs and symptoms which may indicate danger of hearing impairment. A significant outcome of the study will be the development of the methods and techniques for the Subcommittee's national studies.

The initial study, which is planned for a five-year period, is being conducted on a population of approximately 5 000 children ranging in age from three to 15 years. The school-age portion of this population is 97.5% of the enrollment of four public schools. The preschool children are those who are expected eventually to attend the four study schools. The four schools were

TABLE 1. Sound pressure levels (in decibels re 0.0002 dyne/cm² as measured in octave bands listed) recorded during acoustic survey at H. C. Frick School, September 15, 1959. Recording was done both outside and inside the audiometric test room and under varying noise conditions: with hall door closed and open, with test room ventilator on and off, with and without conversation in the room, during lunch period, and during change of classes. Given below also, for purposes of comparison, are the maximum allowable sound pressure levels (SPL) to determine hearing levels at 30 decibels below audiometric zero with no masking.

Octave Bands (in cps)								
20 to 10 000	*20 to 75*	*75 to 150*	*150 to 300*	*300 to 600*	*600 to 1 200*	*1 200 to 2 400*	*2 400 to 4 800*	*4 800 to 10 000*
Outside Test Room								
62	58	54	54	48	42	38	34	28
(hall door closed, ventilator on, conversation)								
58	55	54	48	43	33	36	32	30
(hall door closed, ventilator on, no conversation)								
58	54	52	50	52	46	50	48	30
(hall door closed, ventilator on, lunch hour)								
58	50	55	54	54	51	44	38	36
(hall door open, ventilator off, conversation)								
Inside Test Room, Test Room Door Closed								
43	44	30	11	8	8	9	9	13
(hall door open, ventilator on)								
46	44	24	7	8	8	9	10	13
(hall door closed, ventilator off, classes changing)								
48	49	38	11	9	8	8	9	13
(hall door closed, ventilator off, lunch hour)								
44	46	29	10	8	7	8	9	14
(hall door closed, ventilator off)								
Maximum Allowable SPL								
		10	10	10	10	12	22	32

chosen to be representative, as a group, of the general Pittsburgh public school population from a socioeconomic standpoint.

Individual air-conduction hearing levels are being determined and otological examinations made on the children at varying intervals. Medical histories have been obtained. The relation among otological findings, medical histories, air-conduction hearing levels, academic achievement, and other aspects of child development are being studied.

A prime objective of the Study of Hearing in Children is to obtain information on hearing levels by determining individual air-conduction hearing levels. It was felt that these hearing levels were substantially better than adult norms, and to measure them accurately it was necessary to have both the proper acoustic environment and audiometers which would test with linearity to 30 db below the American Standard audiometric zero. This paper presents the solutions to problems encountered in connection with the acoustic environment and with the audiometers.

Acoustic Environment

The choice of audiometric test rooms was made after an investigation of the various types of rooms commercially available and after a study of their construction and attenuation characteristics. Double-wall rooms were selected because single-wall rooms did not provide sufficient attenuation of ambient noise, particularly for testing at low frequencies.

The audiometric test rooms used in the Pittsburgh Study are prefabricated, portable rooms[1] which were erected in rooms set aside for this purpose in Pittsburgh schools, after acoustic surveys showed that the sites were relatively quiet areas. Each test room is actually a room within a room with an air space between two separate housings; the inner room rests on rubber vibration insulators while the outer room rests on the area floor. Walls and ceilings of both inner and outer rooms are constructed of 4″ thick acoustic panels. The test rooms have two adjacent separately hinged doors that open in opposite directions. Each room contains a double-glazed window, is lighted by an incandescent light and has continuous ventilation. The inside measurements of the rooms are 72 in. by 64 in. Outside measurements are 96 in. by 88 in. The weight is approximately 6 225 lb.

Before the audiometric test rooms were chosen, advice was sought regarding the sound pressure levels which must not be exceeded in the audiometric test rooms if hearing levels of children well below audiometric zero (American Standard) were to be determined without masking. The maximum allowable sound pressure levels in order to test at 30 db below the American Standard audiometric zero (*1*, *2*) are listed by octave band in Table 1.

Acoustic surveys were made at the sites chosen in the four study schools, both before and at periodic intervals after the erection of the audiometric test rooms. In Table 1, the results of a recent survey in one school are summarized. The findings in this survey, which was made after the rooms had been in use over a year, are typical of the conditions found at the remaining three study sites. During the acoustic surveys, sound pressure levels were recorded both inside and outside the audiometric test rooms. The sound pressure levels recorded in Table 1 were the highest noted during test periods which were chosen to include periods of peak noise during the school day. The 150 to 300 cps octave band was of particular concern, since the lowest test frequency was 250 cps. When the attenuating effect of the earphones is added, it can be seen that there is a margin of safety even in the 150 to 300 cps octave band, which allows the determination of hearing levels to 30 db below audiometric zero.[2]

[1]Model 1202, by Industrial Acoustics Co., 341 Jackson Ave., New York 54, N. Y.

[2]Sound level measurements were made by a General Radio Octave Band Analyzer, type number 1550-A, used in conjunction with a General Radio Sound Level Meter, type number 1551-A. Both instruments were calibrated in the Acoustics Laboratory of the Graduate School of Public Health and were found to meet the American Standard Specification for instruments of this type.

Audiometers

Two methods were considered for dealing with the problem of determining hearing levels with linearity below audiometric zero. The first was to construct audiometers which would test to the required levels. Highly accurate laboratory type audiometers could be built, but this would be costly and time consuming. In addition, there was the question of whether a commercially available audiometer could be made to serve the purposes of the study. A plan for building special audiometers was, therefore, not developed. The second method considered, and adopted, was to obtain five commercially available audiometers constructed with auxiliary 40 db precision attenuators placed in the earphone lines. These attenuators could be switched in or out of the circuit and would be switched in on the linear portion of the attenuator scale so that low hearing levels could be measured accurately.

An audiometer manufacturing company, which constructed audiometers as described, guaranteed them to meet the American Standard Specification. A significant feature of the audiometers was their Western Electric 705-A earphones for which the threshold characteristics are known. The audiometers were not required to provide masking or bone conduction. The frequencies supplied were 250, 500, 1 000, 2 000, 4 000, 6 000, and 8 000 cps. The hearing level dials were graduated in 5 db steps from zero to 60 db. The audiometers were of a type that is widely used throughout the country and are produced by a company that cooperated to the fullest extent in complying with its guarantee that the audiometers meet the American Standard Specification. It is emphasized that the experience with audiometer performance, as described on the following pages, is not unique to the make of audiometer used in the study.

Audiometer Performance

Performance at Initial Calibration. First attempts to check the calibration data supplied by the manufacturer against the physical performance of the audiometers were unsuccessful. At this time, the Acoustics Laboratory at the University of Pittsburgh was in the process of development and not able to calibrate the audiometers. A portable artificial ear, which is available commercially, was used instead. The artificial ear was found to have an unstable microphone and not to meet the American Standard Specification. Advice and help was then obtained from the Sound Section of the National Bureau of Standards.

The audiometers failed to meet the American Standard Specification in the following aspects:

> interrupter click on termination of tone; acoustic radiation from audiometer chassis; intensity and frequency changes at onset and termination of signal tone; extraneous signals at certain frequencies; erratic earphone response; frequencies outside the 5% tolerance of the American Standard Specification; rms sound pressure errors of certain earphones greater than the tolerance limits of 4.0 db at the test frequency of 2 000 cps and lower, and 5.0 db at the test frequencies above 2 000 cps; variation in the sound pressure output greater than 2 db as specified in the American Standard Specification when line voltage was varied from 105 to 125 volts; tone interrupter showing rise and decay times of tones outside of specifications; overshoot of tone outside specifications; attentuator linearity

outside the tolerance limit of plus or minus 1.5 db at 10 to 5, and 5 to 0 steps; and tests of purity of test tones showing the sound pressure levels of the fundamental to be less than 25 db above any harmonic at certain frequencies.

Not all of these faults were found in each audiometer. However, none of the five met all specifications at the first calibration, and were returned to the manufacturer for repair or correction or for replacement of unstable earphones. After correction, the audiometers were sent to the Bureau of Standards for their second calibration. At this time, one audiometer was found to be within specifications and went into use in the study in Pittsburgh. The remaining four audiometers were again returned to the manufacturer for further changes and then were sent to the Bureau of Standards for the third calibration. Three audiometers met specifications at the third calibration. The fifth audiometer was again returned to the manufacturer for correction and was finally received in June 1958 and was found to meet the American Standard Specification. A sixth audiometer was purchased later as a spare.

The difficulties experienced in getting the audiometers to meet the American Standard Specification involved a period of five months, and demonstrated the need for comprehensive calibration and repair facilities which would be easily and quickly available. The Department of Occupational Health of the Graduate School of Public Health expanded its Acoustics Laboratory so that by the time the audiometers were ready for use the laboratory was fully equipped to carry out artificial ear calibration of audiometers. Periodic calibration of the hearing study audiometers was then established.

Performance Following Initial Calibration. The audiometers were calibrated as often as possible until sufficient information was available to judge their stability. Beginning in July 1958, 45 complete calibrations were done on six audiometers through July, August, and September. During this period, two of the audiometers were out of operation for several weeks at a time while unstable earphones were replaced. Therefore, an average of three calibrations per month was performed on each audiometer in operation during the three months. In October the interval between calibrations was lengthened to two weeks, and in November to one month. As a result of this experience, the interval between calibrations was set not to exceed one month unless the performance of the audiometer indicated the need for more frequent calibration. Necessary servicing of the audiometers was carried out at the time of calibration so that the American Standard Specification was consistently maintained at all times when the audiometers were in use.

In addition to the monthly acoustic calibration, the audiometers received daily checks by the technicians. Beginning in January 1959, a weekly measurement of the voltage output of the audiometers at the terminals of the earphones was instituted. This measurement is carried out at the site where the audiometer is used. A series of measurements established a baseline of output readings for each frequency. Variation of plus or minus 2 db from this baseline on two successive days was taken to indicate the need for an acoustic calibration.

Problems in Tone Presentation. The most serious problem in performance of the audiometers was delay in rise and decay times of the tone presentation, and overshoot and undershoot of tone. The difficulty arose from the fact that the tone in the audiometers was interrupted by the starting and stopping of a Wien bridge oscillator. All of the audiometers met the American Standard Specification at the start of the study but were close to the tolerance limit in respect to rise time of tone presentation. Overshoot of tone beyond the required plus or minus 1 db of the required sound pressure was sufficient to place one audiometer out of operation early in the study. The remaining audiometers showed overshoot which tended to increase with the passage of time.

The problem of delay in tone presentation, and of overshoot and undershoot was finally solved by the staff

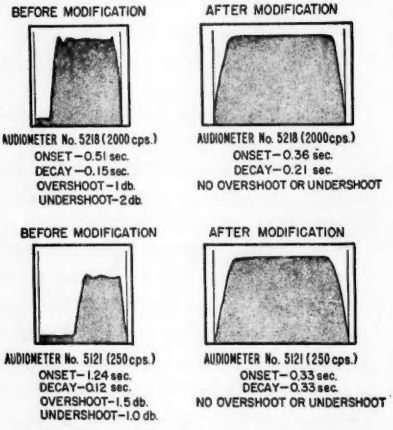

FIGURE 1. Tracings of tone representation in audiometers before and after modification with specially constructed electronic switches to control tone presentation.

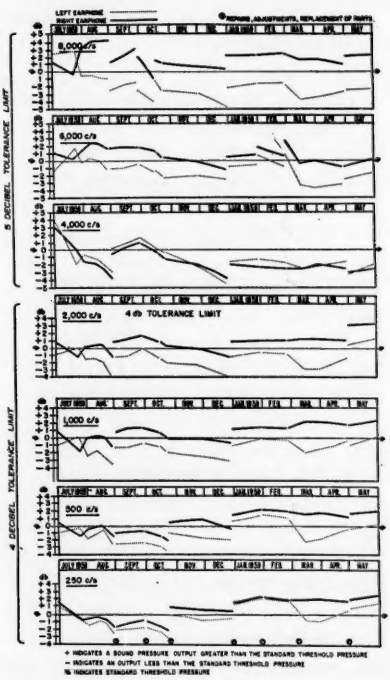

FIGURE 2. Sound pressure errors determined during acoustic calibrations of audiometer number 5 211, showing variation from standard threshold pressures, by earphone, by frequency, and by time. Each break in the output lines indicates one of the occasions on which readjustment or repair was made at one or more frequencies.

of the Acoustics Laboratory. To correct the difficulty, the audiometers were modified by introducing tone interruption through a separate electronic switch placed between the oscillator and the output. By this method the oscillator is left on all the time that the audiometer is turned on and interruption is accomplished by the electronic switch. The need to operate the oscillator continuously introduced a number of electronic problems which

did not exist when the tone was interrupted by stopping the oscillator. These problems were associated with a strong leak-through of tone due to stray wiring capacitance, and with the location of the audiometer components. This last was eliminated only after modification of the audiometer circuits to the extent that the output transformer was removed from its location and mounted elsewhere on the chassis.

These changes have resulted in a smooth tone where the duration of onset is approximately 0.3 sec. Furthermore, there is no discernible overshoot or undershoot. Tracings of the tone presentation before and after tone modification are shown in Figure 1.

Other Major Defects. The major defects in performance of the six audiometers, occurring during the 21-month period from July 1, 1958, to March 31, 1960, in addition to that of tone presentation, are listed below:

Defect in Performance	*Number of Times of Occurrence*
Sound Pressure Output Approaching or Exceeding Tolerance Limits	45
Faulty Earphone Performance	10
Earphones Replaced Due to Performance	6
Frequency Outside 5% Tolerance	8
Harmonic Distortion	7
Extraneous Instrument Noise	4
Worn Interrupter Switch	2
Defect in 40 db Attenuator	4
Defect in Electronic Switch	1

From this tabulation it can be seen that the most common defect in performance was the tendency of the audiometers to produce sound pressure errors greater than the tolerance limits of 4.0 db at the test frequency of 2 000 cps and lower, and 5.0 db at the frequencies above 2 000 cps. This defect occurred 45 times in all six audiometers. The smallest number of occasions on which it occurred in any one audiometer was three times, and the greatest, 13 times.

The sound pressure output of one audiometer is illustrated in Figure 2, for the period July 1, 1958, to May 30, 1959, during which time it was necessary to readjust the audiometer on seven occasions. In the figure, each break in the lines representing the output of the earphones indicates one of the occasions on which readjustment or repair was made at one or more frequencies. In this connection, it should be mentioned that the audiometers were brought into specifications when the sound pressure errors drifted close to the tolerance limits. It is not known how far they would have drifted outside the tolerance limits if the audiometers had not been readjusted.

Not all audiometers behaved in the same manner in respect to output. During the 21-month period previously mentioned, the longest continuous periods each audiometer remained within specifications were: one for eight months, one for five months, one for four months, two for three months, and the sixth audiometer stayed within the specification at no time longer than two months.

Faulty earphone performance necessitating the replacement of earphones on six occasions was the next most frequent problem. On eight occasions, frequencies were outside the 5% tolerance in two of the audiometers. Exces-

TABLE 2. Combined average errors from the 40 db attenuator, the 5 db attenuator steps, and sound pressure output for each earphone for the period March 16 to 30, 1959, for audiometer #5 121.

Hearing Level (in db)	*Frequency*						
	250	*500*	*1000*	*2000*	*4000*	*6000*	*8000*
Right Earphone							
–20	4.6	4.9	6.6	2.0	5.6	4.4	1.9
–15	4.6	4.9	6.6	2.0	5.6	4.4	1.9
–10	4.2	4.5	6.2	1.6	5.2	4.0	1.5
– 5	4.4	4.7	6.4	1.8	5.4	4.2	1.7
0	4.0	4.4	6.1	1.5	5.0	3.8	1.4
5	4.2	4.5	6.2	1.6	5.2	4.0	1.5
10	3.8	4.2	6.0	1.4	4.8	3.6	1.2
15	3.8	4.2	5.9	1.3	4.8	3.6	1.2
20	3.2	3.5	5.2	0.6	4.2	3.0	0.5
25	3.0	3.4	5.1	0.5	4.0	2.8	0.4
30	2.6	3.0	4.7	0.1	3.6	2.4	0.0
35	2.8	3.2	4.9	0.3	3.8	2.6	0.2
40	2.4	2.8	4.6	0.0	3.4	2.2	–0.2
45	2.6	3.0	4.7	0.1	3.6	2.4	0.0
50	2.3	2.6	4.4	–0.2	3.3	2.1	–0.4
55	2.2	2.6	4.4	–0.2	3.2	2.0	–0.4
60	1.6	2.0	3.7	–0.9	2.6	1.4	–1.0
Left Earphone							
–20	0.6	1.4	3.4	0.0	4.4	3.5	–1.2
–15	0.6	1.4	3.4	0.0	4.4	3.5	–1.2
–10	0.2	1.0	3.0	–0.4	4.0	3.1	–1.6
– 5	0.4	1.2	3.2	–0.2	4.2	3.3	–1.4
0	0.1	0.9	2.8	–0.6	3.8	3.0	–1.7
5	0.2	1.0	3.0	–0.4	4.0	3.1	–1.6
10	0.0	0.8	2.6	–0.7	3.6	2.8	–1.8
15	–0.1	0.7	2.6	–0.8	3.6	2.8	–1.9
20	–0.8	0.0	2.0	–1.4	3.0	2.8	–2.6
25	–0.9	–0.1	1.8	–1.6	2.8	2.1	–2.7
30	–1.3	–0.5	1.4	–2.0	2.4	1.6	–3.1
35	–1.1	–0.3	1.6	–1.8	2.6	1.8	–2.9
40	–1.4	–0.6	1.2	–2.1	2.2	1.4	–3.2
45	–1.3	–0.5	1.4	–2.0	2.4	1.6	–3.1
50	–1.6	–0.8	1.1	–2.2	2.1	1.2	–3.4
55	–1.6	–0.8	–1.0	–2.3	2.0	1.2	–3.4
60	–2.3	–1.5	0.4	–3.0	1.4	0.6	–4.1

sive harmonic distortion occurred on one occasion in each audiometer and on two occasions in one audiometer at 500 and 8 000 cps. These occasions followed closely after the previously described modification of the audiometers to correct the defects in tone presentation.

The audiometers remained relatively stable in respect to the 5 db hearing loss intervals or attenuator steps, varying not more than plus or minus 1.5 db at each interval. The cumulative tendency in each audiometer was for the intervals to be less than the 5 db step so that the range from 0 to 60 db was actually not 60 db but usually closer to 55 db.

The 40 db attenuators also remained relatively constant, although each of the six actually attenuated 39 db or slightly less. With the use of these attenuators switched into the circuit at an attenuator dial reading of zero, the audiometers had a range of attenuation from 60 db above audiometric zero to 40 db below, as far as dial settings were concerned, but actually the ranges were 6 to 8 db less than this.

In summary, the following audiometer components produced variations or errors which influenced the hearing levels recorded at each threshold determination: (a) the 40 db attenuators in the earphone lines, a relatively constant variation; (b) the 5 db attenuator steps, relatively constant variations; and (c) the sound pressure output at the various frequencies, errors which varied with time, frequency, earphone, and with audiometer.

Following are examples of the magnitude of errors in these three components for two adjacent calibrations in one audiometer: the average error in the 40 db attenuator equalled 1.6 db; the average error at the 5 db attenuator steps varied at each respective step and ranged from 0.0 db to 1.4 db; the sound pressure output showed average errors by frequency ranging from —1.0 db to +3.8 db in the right earphone and from —4.1 db to +1.4 db in the left earphone.

When these three variations or errors were added algebraically to the results obtained in hearing level determina-

TABLE 3. Difference (in db re audiometric zero) between corrected and uncorrected mean hearing levels (mean hearing levels based upon results of all reliable audiometric tests conducted during the given time period) for right and left ears at selected frequencies for each audiometer used between June 1958 and June 1959. 'Uncorrected' hearing levels are those obtained from the dial readings of the 5 db attenuator in the normal course of testing. 'Corrected' hearing levels are obtained by adding, to the 'uncorrected' result, the sum of the errors inherent in the output of the earphone, the 5 db attenuator steps, and the 40 db attenuator.

Audiometer	*Ears*	*Number of Ears Tested*				*Corrected, Uncorrected Diff*			
		250 cps	*1000 cps*	*4000 cps*	*8000 cps*	*250 cps*	*1000 cps*	*4000 cps*	*8000 cps*
All	Right	6340	6740	6739	6285	2.3	1.8	2.1	1.8
	Left	6338	6741	6739	6285	3.2	2.8	2.2	2.0
#5 040	Right	623	671	671	615	1.0	1.0	1.3	0.5
	Left	624	671	671	615	1.8	3.4	1.9	0.5
#5 121	Right	1011	1058	1056	1003	3.9	2.7	4.2	4.0
	Left	1011	1058	1057	1003	3.7	3.6	2.2	0.7
#5 148	Right	1159	1300	1300	1147	1.7	1.7	3.1	0.7
	Left	1157	1300	1299	1147	3.6	3.3	2.7	4.0
#5 211	Right	1193	1218	1219	1191	3.9	3.9	0.7	4.5
	Left	1193	1219	1219	1191	2.3	3.5	1.3	0.3
#5 218	Right	1063	1152	1152	1053	1.3	1.3	1.8	0.9
	Left	1062	1152	1152	1053	2.6	2.4	2.8	3.5
#5 226	Right	1291	1341	1341	1276	1.6	0.1	1.2	—0.2
	Left	1291	1341	1341	1276	4.1	1.0	2.4	2.4

tions, it was found that in some instances as much as a 5 or 10 db shift was obtained in a reading at a particular frequency. In other instances, no change occurred in a recorded reading. Table 2 is presented to show an example of the magnitude and direction of the combined errors when the average errors in the 40 db attenuator, attenuator (hearing-level dial) steps, and the sound pressure output of each earphone are added algebraically. The variations by hearing level, frequency, and by earphone, in this instance, have a range of approximately 10 db.

Correction of Readings

Because of the magnitude of error as shown in Table 2, the readings obtained in hearing level determinations were corrected first by the appropriate exact numerical error. However, this involved adding an 'exact' error to a relatively gross measurement of hearing level obtained in 5 db steps. Thus, individuals who hear a particular frequency at 10 db and who do not hear at 5 db are recorded as having thresholds of 10 db but actually they may be able to hear at the 6, 7, 8, or 9 db level. If it is assumed that of all individuals with hearing levels recorded as 10 db an equal number begins to hear at each point between 5 and 10 db, then the whole group, although recorded as having a hearing level of 10 db, actually had an average hearing level of 7.5 db. Thus the hearing levels, as determined, may underestimate the hearing levels for the group by 2.5 db. Because of this, it was felt that the hearing levels could be made more exact, although still approximations, by

TABLE 4. Difference (in db re audiometric zero) between corrected and uncorrected mean hearing levels (mean hearing levels are based upon the results of all reliable audiometric tests conducted during the given time period) for right ears at selected frequencies for each month between June 1958 and June 1959. 'Uncorrected' hearing levels are those obtained from the dial readings of the 5 db attenuator in the normal course of testing. 'Corrected' hearing levels are obtained by adding, to the 'uncorrected' results, the sum of the errors inherent in the output of the earphone, the 5 db attenuator steps, and the 40 db attenuator.

Period of Time	*Number of Ears Tested*				*Corrected, Uncorrected Diff*			
	250 cps	*1000 cps*	*4000 cps*	*8000 cps*	*250 cps*	*1000 cps*	*4000 cps*	*8000 cps*
June 1958 to June 1959	6340	6740	6739	6285	2.3	1.8	2.1	1.8
June 1958	390	470	470	390	2.5	1.3	3.4	5.9
July 1958	443	587	587	440	1.6	0.8	2.6	4.5
August 1958	514	640	640	513	1.2	0.6	1.2	3.6
September 1958	389	394	394	388	1.1	1.5	1.7	1.8
October 1958	408	408	408	408	1.6	1.9	2.5	1.5
November 1958	367	368	368	367	1.4	1.1	0.8	0.7
December 1958	561	561	561	561	1.4	1.2	0.2	0.3
January 1959	547	561	561	532	2.8	2.3	0.4	2.2
February 1959	614	618	618	600	2.1	1.5	0.8	1.2
March 1959	302	304	303	300	1.6	2.2	1.9	−0.8
April 1959	767	778	779	761	2.4	2.4	2.4	1.0
May 1959	636	642	641	626	4.6	3.4	4.9	0.8
June 1959	402	409	409	399	4.8	3.7	3.7	1.2

TABLE 5. Difference (in db re audiometric zero) between corrected and uncorrected mean hearing levels (mean hearing levels are based upon the results of all reliable audiometric tests conducted during the given time period) for left ears at selected frequencies for each month between June 1958 and June 1959. 'Uncorrected' hearing levels are those obtained from the dial readings of the 5 db attenuator in the normal course of testing. 'Corrected' hearing levels are obtained by adding, to the 'uncorrected' results, the sum of the errors inherent in the output of the earphone, the 5 db attenuator steps, and the 40 db attenuator.

Period of Time	*Number of Ears Tested*				*Corrected, Uncorrected Diff*			
	250 cps	*1000 cps*	*4000 cps*	*8000 cps*	*250 cps*	*1000 cps*	*4000 cps*	*8000 cps*
June 1958-June 1959	6338	6741	6739	6285	3.2	2.8	2.2	2.0
June 1958	390	470	469	389	2.6	3.6	3.4	5.5
July 1958	443	587	587	441	3.1	3.6	2.4	4.4
August 1958	514	640	640	513	3.6	3.9	1.8	4.1
September 1958	389	394	394	388	3.5	2.6	2.7	3.0
October 1958	408	408	408	408	4.2	3.1	2.7	2.3
November 1958	367	368	368	367	4.2	3.8	1.1	2.5
December 1958	561	561	561	561	4.9	2.4	0.0	0.7
January 1959	546	561	561	532	3.5	3.2	0.4	1.1
February 1959	614	618	618	600	3.7	3.5	1.0	1.0
March 1959	302	304	303	300	3.1	2.3	2.4	−0.2
April 1959	767	779	779	761	3.3	2.4	2.8	0.6
May 1959	635	642	642	626	1.0	0.9	4.7	1.3
June 1959	402	409	409	399	0.6	0.4	4.3	2.0

substracting 2.5 db from each hearing level determination. With this second adjustment, the addition of the exact numerical error as described above becomes more realistic. A further adjustment was the rounding off of the resultant hearing level value to the nearest whole number.

Tables of adjusted and corrected values were computed for each month of the study, for both earphones of each audiometer, by frequency, and for each possible attenuator dial reading. The computation of adjusted and corrected values for a particular month is made by averaging the total error obtained from the calibration at the start of that month with the error obtained at the start of the following month. In those months in which there were calibrations in addition to the monthly calibration, a weighted average is calculated. In computing this average, adjacent calibration results are averaged and weighted by the number of days between the calibrations.

Tables 3, 4, and 5 are presented to show the differences between the readings obtained by correcting the hearing levels as described, and what the readings would have been if they had remained uncorrected. These differences in mean hearing levels have been computed for selected frequencies of 250, 1 000, 4 000, and 8 000 cps.

In Table 3, for the right ear it can be seen that the differences in hearing levels measured by audiometer number 5 040 range by frequency from 0.5 to 1.3 db, and by audiometer number 5 211 from 0.7 to 4.5 db. The differences, therefore, vary in magnitude from audiometer to audiometer. It is

TABLE 6. Mean and median hearing levels (HL), in db re audiometric zero, and standard deviations for children in otologically 'normal' and 'abnormal' populations tested between June 1958 and June 1959.

Frequency (cps)	*Number Tested*		*Mean HL*		*Median HL*		*SD*	
	Norm	*Abnorm*	*Norm*	*Abnorm*	*Norm*	*Abnorm*	*Norm*	*Abnorm*
Right Ear								
250	2 029	412	− 9.8	−5.2	−10.6	−6.9	7.7	12.2
500	2 175	457	− 7.2	−2.2	− 7.4	−4.5	7.6	12.3
1 000	2 175	457	− 5.3	−0.3	− 5.8	−2.7	7.6	12.7
2 000	2 174	456	− 4.4	−0.1	− 4.7	−2.9	8.0	13.0
4 000	2 026	411	− 3.6	+2.3	− 4.1	−1.4	9.2	15.2
6 000	2 040	420	− 1.8	+4.4	− 3.0	+0.7	10.1	15.5
8 000	2 017	407	− 3.0	+3.4	− 3.9	0.0	11.0	17.8
Left Ear								
250	2 029	413	−10.6	−6.8	−11.1	−8.9	7.1	10.7
500	2 175	457	− 7.4	−3.0	− 7.8	−4.8	7.3	11.4
1 000	2 175	457	− 5.2	−0.2	− 5.6	−2.8	7.4	12.7
2 000	2 175	456	− 4.5	−0.7	− 4.8	−2.8	8.2	12.6
4 000	2 026	410	− 3.1	+1.2	− 3.7	−2.2	9.1	12.0
6 000	2 040	419	− 1.4	+3.7	− 2.4	+0.7	10.3	15.7
8 000	2 018	406	− 3.0	+3.0	− 3.8	−0.5	11.1	16.7

also seen that there is an additional variation between right and left ears by audiometer and by frequency.

In Tables 4 and 5, it can be seen that the differences between corrected and uncorrected mean hearing levels vary from month to month. For example, the differences in hearing levels in the right ear range at 250 cps from + 1.1 to + 4.8 db; at 1 000 cps from + 0.8 to + 3.7 db; at 4 000 cps from + 0.2 to + 4.9 db, and at 8 000 cps from − 0.8 to + 5.9 db. Table 5 shows similar ranges of differences for the left ear and again there is seen a variation between right and left ears.

For research purposes it was important to adjust the hearing level determinations in this study, despite the relatively small differences in some instances. When attempting to obtain normative data a shift of a few decibels will change a child from one category of hearing to another at certain hearing levels, and when such a shift is applied generally it will affect thousands of children.

Discussion

As has been stated earlier, it was felt that hearing levels in children are substantially better than adult norms, and to measure them accurately it is necessary to have both the proper acoustic environment and audiometers which will test with linearity to 30 db below the American Standard audiometric zero. To demonstrate that this is so and that children in this study had hearing levels measured well below audiometric zero without masking, the findings on hearing levels determined between June 1, 1958, and June 30, 1959, are summarized in the following paragraphs. The detailed findings are being published elsewhere (*3*).

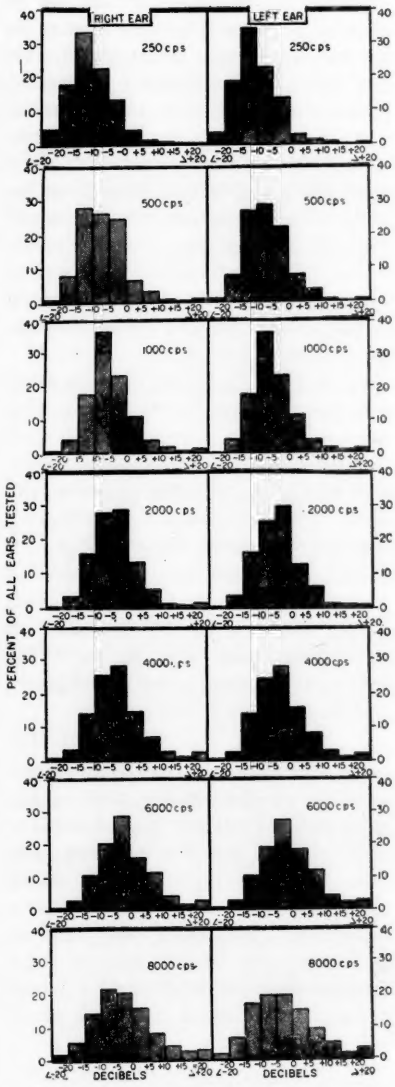

FIGURE 3. Percentage distribution of hearing levels (in db re American Standard audiometric zero) of 2 175 children in an otologically 'normal' population tested between June 1958 and June 1959, each frequency by ear.

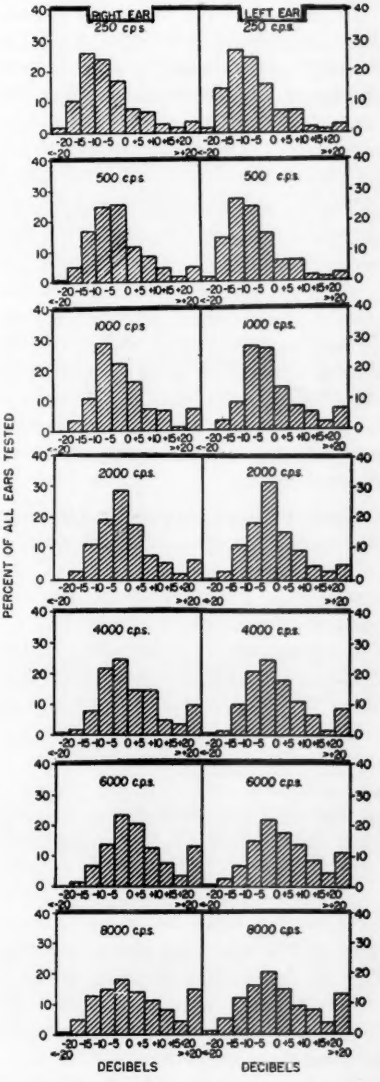

FIGURE 4. Percentage distribution of hearing levels (in db re American Standard audiometric zero) of 457 children in an otologically 'abnormal' population tested between June 1958 and June 1959, each frequency by ear.

During the above mentioned period, 2 175 children, evenly distributed between 5 and 14 years of age, had no otological abnormalities on physical examination. Those children with demonstrable otological abnormalities numbered 457.

The mean and median hearing levels, and the standard deviations for the two groups of children, are shown in Table 6. Both the means and medians of the otologically abnormal group are higher, or less sensitive, than those for the normal children. It is to be noted that these measures of central tendency in both groups of children are significantly lower than the audiometric zero in current use.

Figures 3 and 4 are presented to show the percentage distribution of hearing levels by ear and frequency for the otologically normal and the otologically abnormal groups of children. Of the otologically normal group, 92% had hearing levels lower or more sensitive than the American Standard audiometric zero at 250 cps. This percentage falls progressively as frequencies increase, until at 8 000 cps 65% of the children without abnormalities have hearing levels below audiometric zero. Of 457 children with demonstrable otological abnormalities, 80% have hearing levels below audiometric zero at 250 cps. This percentage also falls as frequencies increase, to 50% at 8 000 cps. These data indicate the high percentages of children with otological abnormalities as well as those without these defects who have hearing levels substantially better than audiometric zero. A study of the distribution in Figure 3 indicates no evidence of masking.

There can be no question as to whether or not audiometers need to be properly maintained within the American Standard Specification if hearing measurement with these instruments is to be done with accuracy. This study demonstrated that at least one commercially available audiometer can be made to meet the American Standard Specification and that it can be modified to test with linearity below audiometric zero, but to accomplish this, proper calibration and repair facilities are needed.

Information obtained from reputable acoustics laboratories in the United States indicates that the experience with audiometers and audiometer performance in this study is not unique to the make of audiometer used. Recent studies (*4, 5*) on the output and general condition of audiometers in use in clinical situations in Scotland provide further evidence of the instability of these instruments and the need for better construction, better maintenance, and more frequent calibration.

A distinction should be made between the steps taken for the research purposes of this study, such as the adjustment of hearing level determination readings, and the necessary precautions taken in regard to maintenance of acoustic environment and audiometer calibration. These latter must be made an integral part of any hearing measurement program whether of a school population, a clinical case load, or an industrial program.

Summary

Preliminary to a long-term, nationwide study of children's hearing problems, an experiment was conducted

in acoustic environment control and audiometer modification and performance. It is concluded that rigid criteria for control of acoustic environment can be met in the field; that audiometers can be modified to test accurately levels well below the American Standard audiometric zero; that precautions taken to control acoustic environment and to check audiometer calibration before and during use should be an integral part of any measurement hearing program.

Mean and median values of hearing levels of children in this study vary with frequency and are more sensitive than the American Standard audiometric zero. A substantial number of otoscopically abnormal ears do not show hearing loss and conversely many ears with hearing loss show no observable otologic abnormality. It is concluded that audiometric testing alone cannot identify physical abnormalities of the ear which may have predictive value.

Acknowledgement

Consultants to the study from the University of Pittsburgh include: Isidore Altman, Ph.D., Professor of Medical Care Statistics; Raymond E. Jordan, M.D., Clinical Professor of Otology; Samuel M. Wishik, M.D., M. P. H., Professor of Maternal and Child Health. The excellent work of the staff of the study is gratefully acknowledged, in particular that of Herbert S. Levine, M.Sc. (Hyg.), Biostatistician; William Melnick, Ph.D., Audiologist; and Mrs. Shirley J. Bakker, Administrative Assistant. The assistance of the Department of Occupational Health in the University of Pittsburgh's Graduate School of Public Health, and in particular of Assistant Professor Kenneth Stewart, Director of its Acoustic Laboratory, is also gratefully acknowlegeded.

References

1. American Standards Association, *American Standard Criteria for Background Noise in Audiometer Rooms*, S 3.1-1960. New York: Am. Standards Ass.
2. Cox, J. R., Jr., How quiet must it be to measure normal hearing? *Noise Control*, 1, 1, 1955, 25-29.
3. Eagles, E. L., and Wishik, S. M., A study of hearing in children: objectives and preliminary findings. *Trans. Amer. Acad. Ophthal. Otolaryngol.*, May-June, 1961.
4. Hinchcliffe, R., and Littler, T. S., Methodology of air-conduction audiometry for hearing surveys. *Ann. occup. Hyg.*, 1, 1958, 114-133.
5. Lenihan, J. M. A., *Scottish Audiometer Calibration Service Annual Report*. Glasgow: Dept. Health Scotland, West. Regional Hosp. Bd., 1957.

Intensive Differential Sensitivity at Masked Threshold

ARNOLD M. SMALL, JR.

FRED D. MINIFIE

Riesz (*14*), in a classical study, 33 years ago reported on the just-noticeable-difference (JND) in intensity for pure tones. He found in general that the JND diminished in size as the sensation level (SL) of the stimulus was increased.

Inasmuch as Riesz's data were gathered under quiet conditions it is pertinent to ask what would happen if JNDs were determined in the presence of a masking stimulus. Sherrick (*16*), with three listeners, has recently investigated this problem using three frequencies, three SLs, and five levels of masking noise. His primary findings indicate that with a method similar to that of Riesz the JNDs are influenced only by the signal-to-noise (S/N) ratio. That is, contrary to Riesz's finding, it is unimportant at what SL the stimulus is presented; the relevant parameter is the ratio of signal to masking noise. As the S/N ratio is increased, the JND becomes smaller. This result is in accord with that reported by Tonndorf, Brogan, and Washburn (*19*).

It has been pointed out that, by using a Békésy audiometry technique, it is possible to estimate JNDs at threshold[1] (*1*). The purpose of the present study was to extend the Riesz and Sherrick information by obtaining with the Békésy technique JNDs for pure tones at threshold in the presence of a masking noise.

Method

Apparatus. Figure 1 shows a schematic block diagram of the apparatus. JNDs at threshold were investigated by means of a motorized attenuator which was part of a Békésy audiometer (Grason-Stadler E-800). Wide-band thermal noise derived from the same source was used as the masker. The signal from a beat frequency oscillator (General Radio 1304B) was led through the motorized attenuator, and signal and masker were mixed in a transformer

Arnold M. Small, Jr., (Ph.D., University of Wisconsin, 1954) is an Associate Professor, Department of Speech Pathology and Audiology and Department of Psychology, University of Iowa. Fred D. Minifie (B.A., Linfield College, 1958) is a Research Assistant, Department of Speech Pathology and Audiology, University of Iowa. This research was supported in part by National Science Foundation Grant G10606.

[1]Hirsh, Palva, and Goodman (*8*) argue threshold swing does not represent differential sensitivity, but variability of the absolute threshold. In spite of this argument it is clear in the present study (see *Discussion*) as well as in others (*1, 13*) that threshold swing bears a marked relation to differential sensitivity as measured just above threshold.

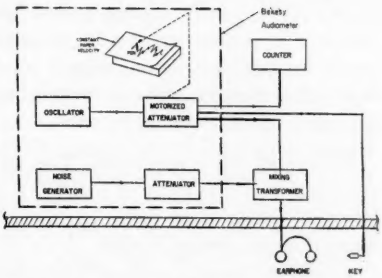

FIGURE 1. Schematic block diagram of apparatus.

(United Transformer, LS-30X). The signal and masker were then led to a single earphone (Telephonic TDH-39 mounted in a MX-41/AR cushion) located in an adjacent room. The active earphone was placed on the right ear of all listeners and the contralateral ear covered with a dummy earphone and cushion. The listener controlled the level of the signal using a cable-mounted, push-button switch. This switch controlled whether the attenuator increased or decreased at a constant rate. In addition, each change in direction was recorded on an electromechanical counter.

Listeners. Each of the 10 listeners was enrolled in an undergraduate course in audiology at the University of Iowa. Even though all listeners had normal hearing as determined by standard air-conduction audiometry (no loss greater than 10 db at any frequency from 125 to 8 000 cps), they varied considerably in listening experience.

Procedure. Each listener was seated comfortably in a dental chair and read a set of instructions. The instructions consisted of (a) a description and illustration of the function of the push-button switch and (b) a detailed account of the listener task. The latter involved listening for the presence of the signal in the background of noise and as soon as the listener could 'just hear the signal' he was to depress the button. As the signal level then decreased, the 'moment' he could no longer hear the signal he was to release the button.[2] This process was repeated continuously for the length of a trial, 3.5 min. It was stressed that the speed of responding when the signal 'appeared or disappeared' was extremely important. In addition, listeners were cautioned not to respond in rhythmic fashion but to base their 'switch activation' on whether they perceived the signal.

TABLE 1. Amount of masking provided by each masker level. Table entries are the difference between the quiet threshold and masked threshold. Thresholds were determined by finding the average midpoint value (of the threshold swings) for the last 3 min of the test run.

Signal Frequency (cps)	*Over-all Masker SPL (db re 0.0002 μbar)*		
	60	*90*	*110*
250	6.2	26.5	44.4
500	11.6	37.5	55.3
1 000	19.2	47.6	67.7
2 000	20.2	48.6	68.1
4 000	20.2	48.8	66.4
8 000	12.3	38.9	56.9

Preliminary tests showed that approximately 15 sec were required for the threshold to stabilize and for the

[2]This procedure, of course, is the standard manner of obtaining thresholds with a Békésy audiometer. It differs from tradition only in that the signal is presented against a constant noise background.

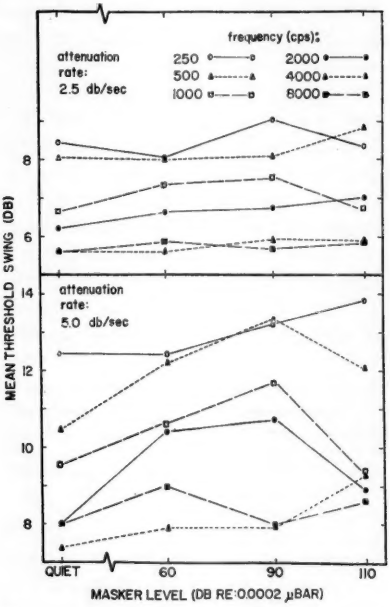

FIGURE 2. Mean threshold swing for 3 min trials averaged over listeners. The upper panel is for the slower (2.5 db/sec) attenuation rate while the lower panel shows the results for the faster (5.0 db/sec) attenuation rate. The parameter within each panel is signal frequency.

listener to get the 'feel' of the situation. It was anticipated, therefore, that 30 sec should provide ample time for threshold stabilization. Consequently, although each trial lasted 3.5 min, data were recorded only during the last 3 min of each trial and not during the first 0.5 min. Each experimental session consisted of 12 such trials separated by 1 min with a 3 min rest period after the sixth trial.

The first listening session consisted of practice. The listeners were presented a variety of stimuli chosen from those to be given during the actual experiment. At the end of the period all listeners appeared to be reasonably facile in manipulation of signal level and quite consistent from listener to listener in their response patterns.

Design of the Experiment. The experiment took the form of a 6 x 4 x 2 factorial design with each of the 10 listeners participating in each of the 48 experimental conditions. In addition the entire experiment was replicated once. Six signal frequencies (250, 500, 1 000, 2 000, 4 000, and 8 000 cps) were presented. Four masker levels (quiet, 60, 90, and 110 db over-all SPL re 0.0002 μbar)

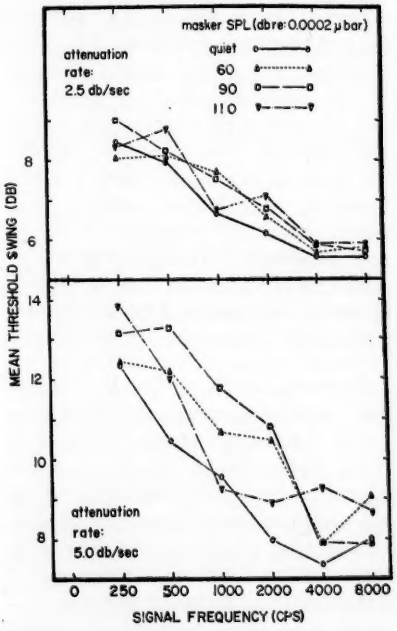

FIGURE 3. Mean threshold swing for 3 min trials averaged over listeners. The upper panel is for the slower (2.5 db/sec) attenuation rate while the lower panel shows the results for the faster (5.0 db/sec) attenuation rate. The parameter within each panel is masker level.

and two attenuation rates (2.5 and 5.0 db/sec) also were used.

In a given session each listener responded under each of the two attenuation rates. With each attenuation rate all six signal frequencies were presented in random order. In alternate sessions the first six trials were given at the fast rate, with the last six presented at the slow rate. One-half of the listeners began with the fast rate; one-half began with the slow rate. During each session constant masker level was maintained. The order in which masker levels were presented in successive sessions was random.

After all listeners had participated in all conditions, the experiment was repeated using the same listeners but with a different randomization of masker level presentation. Different random orders of signal frequency also were used for the replication.

Determination of Differential Sensitivity. It was assumed that by use of the present procedure the average swing in db of the Békésy pen most adequately represented the JND. Rather than reading the amplitude between successive pen reversals from the paper chart, an automatic count of pen reversals was made. In order to determine the average swing during the 3 min test period the simple relation $TS = R_A D/N_R$ was used where TS is mean threshold swing (db), R_A is attenuation rate (db/min), D is duration of test run (min), and N_R is number of pen reversals (counts).

TABLE 2. Analysis of variance of mean threshold swing.

Source	*df*	*ms*	F_1*	F_2†
Frequency (F)	5	7 944.9	46.54**	15.52**
Masker Level (M)	3	831.3	4.87**	
Attenuation Rate (R)	1	38 354.8	224.69**	37.14**
Listeners (L)	9	6 231.6	36.51**	
F x M	15	209.1	1.22	
F x R	5	398.6	2.34	
F x L	45	511.9	3.00**	
M x R	3	297.9	1.75	
M x L	27	973.6	5.70**	
R x L	9	1 032.7	6.05**	
F x M x R	15	124.4		
F x R x L	45	80.4		
M x R x L	27	127.5		
F x M x L	135	170.0		
F x M x R x L	135	93.4		
Within	480	170.7		
Total	959			

*F_1 was calculated by using ms_{within} as the denominator.

†F_2 was calculated by using as the denominator the simple interaction *ms* involving listeners and main effect being tested.

**Significant beyond the 0.01 level.

Results

Table 1 indicates the amount of masking produced under each of the experimental conditions. The data of Table 1 agree with that reported by Hawkins and Stevens (7), in that there is a nearly one to one relation between masker level and amount of masking as long as there is more than about 15 db of masking. For example, if the masker level is increased by 20 db, so also is the amount af masking.

Figure 2 shows the mean threshold swing for each trial under each of the experimental conditions averaged over listeners. For the slower attenuation rate there seems to be no marked effect of masker level for those levels used in the present study. The faster attenuation rate shows greater variability but, again, no systematic trends as a function of masker level. On the other hand, effects of frequency are readily seen in Figure 2. In general, low signal frequencies show relatively larger swings while high frequencies show smaller swings.

The effect of signal frequency is even more apparent in Figure 3 which shows the same data as in Figure 2, but plotted in a different manner. Figure 3 also shows to a better advantage a slight tendency (though not statistically significant, see Table 2) for the mean threshold swings to be proportional to masker level for the three lowest levels at least.

Table 2 shows a summary table of an analysis of variance performed on the mean threshold swings. When the error term is that derived on the basis of the replication of the experiment, all main effects as well as all simple interactions involving listeners are significant beyond the 0.01 level (*4*, pp. 237-247). Thus, within the context of this specific experiment, for those conditions and listeners used, the signal frequency, the masker level, the attenuation rate, and the listeners, plus all simple interactions involving listeners, contributed significantly to the total variance.

It is generally desirable to generalize beyond the specific conditions employed in a given experiment. This can be done readily if at least one variable represents a random sample from some specifiable population. In this case it may be assumed that listeners are a random sample of listeners in general. Then for listeners in general it is found that only signal frequency and attenuation rate contribute significantly to the total variance (*4*, pp. 247-252).

Discussion

The results of the present study then would seem to indicate that masker level had little effect upon threshold swing, that is, threshold differential sensitivity. In view of the previously reported relation (*16*, *19*) between S/N ratio and JNDs it would seem pertinent to examine S/N ratio at masked threshold.

It is often assumed that at masked threshold the energy in the critical band is exactly equal to the energy of the signal (7). This is the equivalent of saying that the S/N ratio within the critical band is constant for all masked thresholds, that is, 0 db. The size of the critical band as defined in this manner compares favorably with the size determined by more direct methods (*5*, *15*). This would seem therefore to substantiate the assertion that at masked threshold S/N ratio is constant. If S/N

TABLE 3. A comparison of differential sensitivity. The data of Riesz (*14*) were obtained at 5 db SL by use of a modulation technique. They have been converted from their original form of a just discriminable change in intensity, ΔI, to a just discriminable change in sound pressure, ΔP, so that they would be comparable with ΔPs (threshold swings) obtained in this study. Data from the present study were gathered under the slow attenuation rate with no masking.

Frequency cps	ΔP *(Riesz) db*	ΔP *(Present Study) db*
250	9.0	8.5
500	7.6	7.9
1 000	6.1	6.6
2 000	5.6	6.1
4 000	5.0	5.5
8 000	8.2	5.5

ratio is constant at masked threshold and if S/N ratio is the prime variable influencing the JND, then it is not surprising that in this study the JND was found to be independent of masker level.

It is surprising, if this argument is extended, that the size of the JND changed as a function of frequency. Presumably, S/N ratio within the critical band does not change as a function of frequency. It should be noted, however, that the JND as measured by more traditional psychophysical procedures also changes as a function of frequency. Table 3 compares the data of Riesz (*14*) with data obtained in the present study. The data of Riesz are obtained 5 db above threshold; those from the present study were gathered at the slow attenuation rate under quiet conditions and, of course, at threshold. The results of the two studies agree closely (within 0.5 db) except at 8 000 cps. This agreement seems rather remarkable (perhaps coincidental) in view of the differences in experimental procedures. It might be added once again that although there is substantial agreement for threshold or nearly threshold values, Riesz's JNDs decrease markedly with an increase in signal level while in the present study no significant decrease is seen.

Békésy, as early as 1947 (*1*), suggested that the relative size of the threshold swing might prove to have diagnostic value in differentiating between types of hearing loss.[3] Other investigators have proposed basically similar notions using differential sensitivity at suprathreshold values as their diagnostic index (*3, 10, 11, 12, 13*). The assumption common to all these workers is that certain types of loss exhibit recruitment; that is, an abnormally rapid growth in loudness as stimulus intensity is raised. As a consequence of this rapid growth of loudness, less intensity change is necessary before the change may be discriminated, hence, smaller threshold swings and smaller differential thresholds.

As an aid to better understanding the recruitment phenomena, attempts have been made to simulate recruitment in the laboratory. Auditory fatigue (*2, 6*) and simultaneous masking (*17, 18*) are two methods that have commonly been used. With either of these methods 'recruitment like' effects are seen if a loudness balance technique is used. In addition, it has been reported that, with a Békésy audiometer, a normal ear shows smaller threshold swings after auditory fatigue than before (*1*). How-

[3]Jerger (*9*), as part of a recent article on the clinical application of Békésy audiometry, has provided a rather exhaustive bibliography with 37 citations.

ever, the findings of the present study are contrary to this general similarity between 'real' recruitment and simulated recruitment. Specifically, with simultaneous noise masking no change appears in the size of the threshold swings as the threshold is depressed by increasingly larger amounts of masking noise.[4] On this basis at least, it would seem wise to raise the question as to whether the mechanisms of 'real' recruitment and recruitment simulated by simultaneous masking are indeed identical.

Summary

The purpose of this study was to extend information provided by previous research by obtaining with the Békésy technique JNDs for pure tones at threshold in the presence of a masking noise.

From records made with a Békésy audiometer an analysis was made of the magnitude of the pen swings during threshold determinations. Thresholds were obtained from 10 normal hearing listeners as a function of masker level, signal frequency, and attenuation rate.

Threshold pen swings became larger as either signal frequency was decreased or attenuation rate increased. Masker level had no significant effect.

Summario in Interlingua

Le proposito de iste studio esseva extender information proviste per previe recerca per obtenente con le Békésy technica JNDs pro pur tonos a limine coram un mascante strepito.

De records faceva con un Békésy audiometer un analyse esseva facite de le magnitude de le penna oscillations durante determinations de limine. Limines esseva obtenite de dece ascoltators con normal audition como un function de mascara nivello, signal frequentia, e rata de attenuation.

Limine penna oscillations deveniva plus grande como cata signal frequentia esseva discrescite o rata de attenuation augmentava. Mascara nivello habeva non significant effecto.

References

1. Békésy, G. v., A new audiometer. *Acta Otolaryng.*, 35, 1947, 411-422.
2. Békésy, G. v., Zur Theorie des Hörens; Über die eben markbare Amplitudenund Frequenzänderung eines Tones; Dies Theorie der Schwebungen. *Phys. Z.*, 30, 1929, 721-745. (Reprinted in G. v. Békésy, *Experiments in Hearing*. New York: McGraw-Hill, 1960.)
3. Denes, P., and Naunton, R. E., The clinical detection of auditory recruitment. *J. Laryng.*, 64, 1950, 375-398.
4. Edwards, A. L., *Experimental Design in Psychological Research*. New York: Rinehart, 1950.
5. Fletcher, H., Auditory patterns. *Rev. mod. Physics*, 12, 1940, 47-65.
6. Harris, J. D., Haines, H. L., and Myers, C. K., Loudness perception for pure tones and for speech. *Arch. Otolaryng.*, 55, 1952, 107-133.
7. Hawkins, J. E., Jr., and Stevens, S. S., The masking of pure tones and of speech by white noise. *J. acoust. Soc. Amer.*, 22, 1950, 6-13.
8. Hirsh, I. J., Palva, T., and Goodman, A., Difference limen and recruitment. *Arch. Otolaryng.*, 60, 1954, 525-540.
9. Jerger, J., Bekesy audiometry in the analysis of auditory disorders. *J. Speech Hearing Res.*, 3, 1960, 275-287.
10. Jerger, J., DL difference test, improved method for clinical measurement of recruitment. *Arch. Otolaryng.*, 57, 1953, 490-500.

[4]S. N. Reger (American Speech and Hearing Association meetings, November 1959) reported similar findings with respect to recruitment at threshold with noise masking.

11. Jerger, J., Shedd, Joyce Lassman, and Harford, E., On the detection of extremely small changes in sound intensity. *Arch. Otolaryng.*, 69, 1959, 200-211.
12. Lüscher, E., and Zwislocki, J., A simple method for indirect monaural determination of the recruitment phenomenon. *Acta Otolaryng.*, Suppl. 78, 1948, 156-168.
13. Reger, S. N., A clinical and research version of the Bekesy audiometer. *Laryngoscope*, 62, 1952, 1333-1351.
14. Riesz, R. R., Differential intensity sensitivity of the ear for pure tones. *Phys. Rev.*, 31, 1928, 867-875.
15. Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O., The frequency selectivity of the ear as determined by masking experiments. *J. acoust. Soc. Amer.*, 22, 1950, 490-496.
16. Sherrick, C. E., Jr., Effect of background noise on the auditory intensive difference limen. *J. acoust. Soc. Amer.*, 31, 1959, 239-242.
17. Small, A. M., Jr., and Thurlow, W. R., Loudness relations in two-component tones. *J. acoust. Soc. Amer.*, 26, 1954, 381-388.
18. Steinberg, J. C., and Gardner, M. B., The dependence of hearing impairment on sound intensity. *J. acoust. Soc. Amer.*, 9, 1937, 11-23.
19. Tonndorf, J., Brogan, F. A., and Washburn, D. D., Auditory difference limen of intensity in normal hearing subjects. *Arch. Otolaryng.*, 62, 1955, 292-305.

► *RESEARCH NEWS NOTE*

Cinefluorographic Speech Research

► Speech physiology of cleft palate and normal speakers is being investigated by cinefluorographic and other means at the University of Kansas Medical Center. Goals include development of improved procedures for making measurements from cinefluorographic film; evaluation of movements of subjects' speech mechanisms during phonation of sounds in isolation and in various word and connected speech contexts; and integration of data from articulation testing, voice rating, oral and nasal pressure measurement, and radiographic observation.

The principal item of equipment is a cinefluorographic unit with dynapulse. The nine-inch image intensifier makes possible visualization of the adult speech mechanism from lips to cervical vertebra and from epipharynx to trachea. Pilot work, using a camera speed of 24 frames per second, has shown consistent intrasubject patterns of structural movement for repeated productions of the consonant /r/ in various phonetic contexts, even though, as Moll suggested (*J. Speech Hearing Res.*, 3, 1960, 227-241), certain contacts and rapid movements may be missed at this speed. This concept of consistency of patterns of movements is supported by literature reviewed by Shelton (Ph.D. dissertation, University of Utah, 1959), and by Van Riper and Irwin (*Voice and Articulation*, Prentice-Hall, 1958).

Adequacy of camera speed will be investigated by comparing certain articulatory patterns filmed at 24 and at 64 frames per second. It is suggested that although the former speed appears to give reliable phonetic data, the higher speed may demonstrate finer movements which, though possibly not essential to speech sound production, could prove important to a complete understanding of speech physiology.

To be studied also are procedures for efficient film analysis. This will include the matter of measurements from cinefluorographic film and the problem of choosing landmarks in making such measurements.

The study is sponsored by the National Society for Crippled Children and Adults. Equipment for this and other studies has been provided by the National Institutes of Health, the Kansas Cancer Society, and the Office of Vocational Rehabilitation.

Ralph L. Shelton, Jr., Ph.D.
Assistant Professor
Department of Hearing and Speech
University of Kansas Medical Center

Visual Paired-Associates Task with Deaf and Hearing Children

HANS G. FURTH

Some recent investigators compared the performance of hearing and deaf people on various tasks of visual memory and rationalized their findings in terms of various theoretical notions. Blair (*1*) observed that deaf children were inferior to hearing children on tests such as digit span and thought that those tasks involved a 'relatively abstract type of mental operation.' However, the deaf child's superiority on some other tasks was 'conceived as being the result of a psychological compensatory phenomenon.' Hiskey (*5*), who noticed that deaf children were consistently inferior to hearing children in immediate recall, suggested more simply that 'the hearing children have a decided advantage where vocalization aided retention.' Naffin (7, p. 412), summarizing research of German workers, gave evidence for a marked superiority of the deaf child's visual memory in certain specific situations which, he noted, were characterized by a perceptual poverty or simplicity and required little or no language to be meaningfully perceived. In view of these somewhat conflicting results it would seem that no single generalized attribute can be legitimately affixed to a deaf person's memory.

In search of a unifying theory linking findings on memory with other cognitive areas, it would be important to single out relevant parameters and report results descriptively. Thus Doehring (*2*) limited his investigation to visual spatial memory and chose a task specially designed for this aspect of memory. He used two age groups and consequently could observe performance as a function of age. He found that the performance of deaf children was in every respect similar to that of hearing children.

The present study was concerned with another basic aspect of visual memory, the perception and memory of simple color. A paired-associates task was chosen, involving four simple colors and two toy animals because the experimenter observed that six-year-old deaf children can easily learn to associate the four colors and the two objects with gestures or names. Hence, it was assumed that on this task deaf children were at no disadvantage compared to hearing children as far as language was concerned.

Procedure

Subjects. The deaf group consisted of 180 children of whom 136 came

Hans G. Furth (Ph.D., University of Portland, 1960) is Assistant Professor, and Director, Rehabilitation Training Program, Department of Psychology and Psychiatry, Catholic University of America. The research was supported by a fellowship, MF 9902, from the National Institute of Health, Public Health Service.

from the three schools for the deaf in a western state and 44 from a similar school in a neighboring state. Only three children from the total available were excluded because they failed to meet the criteria of early onset of deafness and a hearing loss of at least 50 db in the better ear. It is believed, therefore, that a reasonably representative sample of deaf children was achieved. Subjects were divided into six age groups (7, 8, 9, 10, 11, 12 years) regardless of class placement: a child whose age was at least 6.5 years and not yet 7.5 on the day of testing, belonged to the 7-year group, and similarly for the other age groups. There were 30 subjects in each of the six age groups. The control group of 180 hearing children was obtained by taking 36 pupils from each of five different grade schools in a city. Each school provided a list of all pupils between the desired ages and by random selection three boys and three girls were chosen from each of the six age groups. The children were evenly divided by sexes in all of the hearing groups and in the 8-, 10-, and 11-year deaf groups. In the 7- and 9-year deaf groups there were 18 boys and 12 girls, and in the 12-year deaf group 17 boys and 13 girls.

Task. The stimulus material consisted of 40 6″ x 6″ colored cards, 10 each of yellow, blue, red, and orange. On the reverse side of each of the orange and yellow cards was a large M; on the reverse side of each of the red and blue cards was a large C.

Method. The experiment was conducted in a familiar room of the school building with only the child and the experimenter present. They were seated across a table from each other. The experimenter handed the child two toy animals, a cat and a mouse. By nonverbal instructions which made use of the letter on the reverse side of the colored cards he explained the correct association of letter and animal: C (red and blue) for cat, M (orange and yellow) for mouse. When the child showed by his immediately successful performance that he associated the letter with the correct animal, the plain (unlettered) side of the 40 colored cards was then shown in the following sequence of

TABLE 1. Mean number of successful performances on memory task (10 correct choices in succession) by 180 hearing and 180 deaf children, by age groups (N = 30 each group), and results of the chi-square test.

Age Groups (in years)	*Performances*		*Chi Square**
	Hearing	*Deaf*	
7	5	7	.39
8	4	5	.00
9	6	8	.09
10	10	8	.08
11	17	7	5.62
12	22	12	5.50
Total	64	47	3.33

*A chi-square value of 3.84 ($df = 1$) is required for significance at the 5% level.

TABLE 2. Mean total of errors, by age groups (N = 30 each group), for 180 hearing and 180 deaf children on memory task (first four of each subject's 40 trials not included).

Group	*Age*						*All Ages Combined*
	7	*8*	*9*	*10*	*11*	*12*	
Hearing	15.0	13.7	12.1	10.9	6.5	6.3	10.76
Deaf	13.9	14.0	11.5	11.4	14.0	12.1	12.81

the four colors (Y, yellow; B, blue; R, red; O, orange): YBRO, YRBO, BYOR, OYBR, BORY, OBRY, ROYB, YORB, RYBO, YRBO. The child's response to the sight of the colored card was to pick up one of the toy animals and place it on top of a small box. Each time he was told whether or not his choice was correct. Each subject received 40 trials, that is, one showing of each of the 40 cards.

A successful performance was defined as 10 correct choices in succession during the course of the 36 trials following the first four trials. The possibility of this performance by chance has been found to be less than .02 (*3*).

Results

Results (Table 1) showed that performances of the corresponding deaf and hearing age groups differed significantly only at 11 and 12 years. The hearing children between the ages of 7 and 12 showed an improvement with age; the chi square for testing the change in relative proportion with age was 24.33, significant beyond the .01 level. In contrast, the corresponding chi square for the deaf group was not significant, 4.43.

The relative absence of change with age for the deaf sample on this memory task is seen even more strikingly in Table 2 where a different computational procedure was used, namely, total number of errors, not counting the first four trials.

An analysis of variance on the data of the foregoing table, summarized in Table 3, revealed significant *F* ratios for the hearing-deafness variable and for the age variable. In view of a significant interaction, group differences at each age level and age differences for each group were tested. A

TABLE 3. Results of variance analysis for testing differences between hearing and deaf groups and among ages on memory test.

Source	*df*	*Variance Estimate*	*F*
Hearing–Deaf	1	378.22	10.81*
Age	5	249.55	7.13*
Interaction	5	197.44	5.64*
Within Cells	348	34.99	
Total	359		

*Associated $p < .01$.

critical difference of 2.99, associated with a .05 level of confidence, established that only the 11- and 12-year-old hearing groups made significantly fewer errors than the corresponding deaf groups and that a striking improvement in performance took place between the 10- and 11-year-old hearing group.

It may be of interest to observe that in both samples girls were consistently superior to boys. In over-all sex comparison 27.8% of the hearing boys had successful performances against 43.3% of hearing girls, and for the deaf sample the corresponding percentages were 24.5% and 32.9%. Chi square for testing the significance of the sex differences was 4.7 ($< .05$) for the hearing group, but only 1.2, not significant, for the deaf group.

In over-all performance, thus, the hearing surpassed the deaf group but this superiority was due to the failure of the 11- and 12-year-old deaf children to match the performance of the hearing children. For the age levels 7 to 10 there was no significant difference between the two groups.

Discussion

The results illustrate that age, or, better, the interaction of age and performance on a particular task, is a relevant variable which must be considered before any generalization concerning the visual memory of deaf children can be made. Between 7 and 10 years deaf children were equal to hearing children of the same age, but 11- and 12-year-old deaf children were decidedly inferior on the same task which was used with the earlier age groups.

By inspection of Table 1, it can be seen that the age at which 50% of a hearing group might be expected to succeed on this task would be between 10 and 11 years while a deaf group might not be expected to reach the 50% level of success within the given age range.

Since only a minority of younger subjects succeeded with 10 successive correct choices, usually only after a considerable number of errors, it is not surprising that the mean total of errors of the age groups 7 and 8 (Table 2) was only barely below the chance level of 18.0 errors in 36 trials. Applying the critical difference of 2.99 as above, it can be asserted that even the poorest learners showed better than chance performance. The finding of an equal memory performance of the younger deaf and hearing children would indicate that the young deaf child's ability for visual color memory is equal to that of the hearing child. If deafness caused some basic change in general perceptual ability one could expect systematic changes at all age levels, but particularly during the early age.

If the foregoing discussion can be construed to mean that the basic capacity for visual memory in deaf children is equal to that of hearing children, some other explanation must be found for the inferiority of the 11- and 12-year-old deaf groups. In this connection it should be pointed out that results similar to the ones obtained on the paired-associates task were found also with concept learning tasks (*3*). Is it possible that the cognitive ability of deaf children around 11 and 12 years old does not keep up with

'normal' development? Or could it be assumed that some motivational factor, extrinsic to cognition as such, may be at work?

It may be suggested that hearing children through training and experience in life and in school gradually develop habits which facilitate learning and improved test attitudes and that on these two counts they gain an advantage over deaf children. Piaget's studies (*6*, p. 337) suggest that normal children begin the process of reflective thought by 11 years and it is in this sense perhaps, as Hiskey (*5*) concludes, that hearing children learn to make better use of their tools for cognition which are mainly words and language.

In contrast to the hearing child, the deaf child between 7 and 10 years of age spends most of his time in school attempting to master the basic elements of ordinary spoken and written language. Due to the pressure of language teaching, relatively little time and emphasis are given to develop and apply the deaf child's cognitive and reasoning abilities, a fact which may adversely influence the motivation and learning attitude of the older deaf child.

Concerning the motivational component it seems noteworthy to report that several teachers in three different schools for the deaf mentioned to the experimenter that their pupils, when they reach the age of starting in the intermediate grades, show a lessening of interest and lack of progress in scholastic achievement. The age period at which this change occurs corresponds to the 11-year-old age group used in this study.

The experimenter also checked the possibility that brighter deaf youngsters might have left schools for the deaf by the age of 11 or 12 to attend hearing schools. It was ascertained that four potential subjects had been omitted from the experimental group for this reason; it is clear that such a small sampling error could not appreciably change the average results. There remains, of course, the final possibility that the lowered performance of the 11- and 12-year-old deaf children was due to a chance factor in the particular state and at the particular time in which the sample was taken.

A trend in scientific evidence during recent years indicates that differences in capacities of the deaf and the hearing children are less than they had been presumed to be. The results of this study tend to support this trend for the area of visual color memory. At the same time these results point to the superiority of the older hearing over the older deaf children and to the different kind of experience and training a deaf child receives. Whether this last factor can be improved upon should be a serious concern for the educator of deaf children.

Summary

A nonverbal memory problem in the form of a visual paired-associates task was given to 180 deaf and 180 hearing children between the ages of 7 and 12 years. Norms of success on this task were established for each age group in both samples. No differences between the deaf and the hearing children were observed at ages 7 to 10. This result tended to

support the theory that the basic capacity of visual memory in deaf children is equal to that of hearing children.

At the 11- and 12-year level the hearing children showed a rapid improvement which was not matched by the deaf children. The deaf children's relative inferiority was thought to be due to a deficiency in experience and training, affecting their cognitive and motivational learning attitude.

Acknowledgment

The writer wishes to thank Dr. Nissim Levy for his assistance and guidance in this research.

References

1. Blair, F. X., A study of the visual memory of deaf and hearing children. *Amer. Ann. Deaf*, 102, 1957, 254-263.
2. Doehring, D. G., Visual spatial memory in aphasic children. *J. Speech Hearing Res.*, 3, 1960, 138-149.
3. Furth, H. G., The influence of language on the development of concept formation in deaf children. *J. abnorm. soc. Psychol.*, in press.
4. Grant, D. A., Additional tables of the probability of 'runs' of correct responses in learning and problem-solving. *Psychol. Bull.*, 44, 1947, 276-279.
5. Hiskey, M. S., A study of the intelligence of deaf and hearing children through a comparison of performances on the separate standardizations of the Nebraska Test of Learning Aptitude. *Amer. Ann. Deaf*, 101, 1956, 329-339.
6. Inhelder, Bärbel, and Piaget, Jean, *The Growth of Logical Thinking from Childhood to Adolescence; an Essay on the Construction of Formal Operational Structures*. New York: Basic Books, 1958.
7. Naffin, P., Die psychologischen Voraussetzungen der Erziehung des taubstummen Kindes. In H. Hetzer (Ed.), *Pädagogische Psychologie*. Göttingen: Hogrefe, 1959.

► RESEARCH NEWS NOTE

Effect of Educational Audiology on Children's Language Development

► A study is planned to investigate whether or not Huizing's 'acoupedic method' (also called 'educational audiology') aids in the development of speech and language of the hearing-impaired child to the extent that he might be integrated into normal hearing classrooms.

An initial sample of 60 hearing-impaired children between the ages of six months and three years will be selected for longitudinal study. Data pertaining to speech and language development will be collected on three levels, representing prelanguage, early speech, and intermediate language development. The Huizing method will be used with this group and at the time each child reaches the age of five, data will be collected for comparison with the normal-hearing child and with the hearing-impaired child trained via more traditional methods. The comparison will be made on vocabulary, social maturity, articulation, and structural analysis of language, on the assumption that predictions pertaining to the possibility of a child's being integrated into a normal hearing classroom might be made prior to his entrance into kindergarten.

The investigation will take five years to complete and is supported by a research contract with the U. S. Office of Education.

J. L. Stewart, Ph. D.
Director, Hearing Center
University of Denver

Lipreading and Letter Prediction

CORINNE M. TATOUL

G. DON DAVIDSON

For some years attempts have been made to isolate variables related to lipreading ability, but to date these variables remain relatively obscure, and the question as to what significant abilities are possessed by good lipreaders is still far from being answered.

To the present writers the most promising variables for further study are synthetic ability, that is, the ability to anticipate the whole from the knowledge of a few of its parts; visual perception, with all that it might include; and attention and concentration, the strong application of one's faculties to the task at hand. Of these, the first variable, synthetic ability, was selected for investigation in this study.

Synthetic ability has long been considered important to successful lipreading by the foremost instructors of the deaf and hard-of-hearing. Nitchie *(4)* emphasized its importance, as did Bruhn (*1*) who believed that the student who tried to see every movement on the lips took longer to learn to lipread than the one who 'let his mind fill in' what he did not see. Howell (*2*) believed that the ability to associate ideas was important and that a person with an 'analytical mind' would have more difficulty in mastering lipreading than would one with a 'synthetic mind.' Simmons (*6*), who most recently studied this variable, suggests the possibility that synthetic ability is necessary and related to lipreading ability, but that a satisfactory means of measuring it is yet to be found.

In the present study synthetic ability was measured by means of scores on a letter prediction test. Letter prediction had been previously investigated by Shannon (*5*) who studied a person's ability to guess the successive letters of a short unfamiliar passage, hypothesizing that anyone speaking a language possesses an enormous knowledge of the statistics of a language. He found that for 129 letters, 69% could be guessed correctly.

Mulder (*3*) studied letter prediction ability using six phrases of 15 letters each, hypothesizing that the more sophisticated a person is concerning language the more of the statistics of the language he knows, and that, therefore, Americans should be able to predict the letters of English more readily than nonnative speakers of English. In general he found this to be true.

It was the purpose of the present study to determine whether relatively

Corinne M. Tatoul (M.S., Florida State University, 1960) is Graduate Assistant, Department of Speech, Florida State University. G. Don Davidson (Ph.D., Ohio State University, 1955) is Acting Associate Professor of Speech Pathology and Audiology, University of Tennessee.

TABLE 1. Means, standard deviations, and ranges of lipreading scores for good and poor lipreaders.

Lip-readers	*Mean*	*SD*	*Range*
Good	133	13	120 to 170
Poor	71	10	50 to 90

good lipreaders, as measured by scores on a lipreading film test, were significantly better than poor lipreaders with respect to synthetic ability as measured by scores on a letter prediction test.

A previous pilot study by one of the authors, using six good and six poor lipreaders and a five-sentence letter prediction test, found the better lipreaders to be significantly better letter predictors.

Procedures

Subjects. The score made on a lipreading film test by 100 college students enrolled in the beginning speech course at Florida State University were ranked and the highest and lowest 25 used to select the experimental subjects (see Table 1). None of the subjects had had any previous lipreading training. Normal vision and hearing were screened with the Snellen Eye Chart and a pure tone audiometer. The criteria for passing were at least 20-20 vision and +15 db hearing acuity, respectively.

Measurement of Lipreading Ability. Lipreading ability was measured by scores on the John Tracy Clinic Film Test of Lip Reading, Form A (7), a silent motion picture test in color on 16 mm film. The test consisted of 30 unrelated sentences which were simple in content and might have been used in everyday life. The sentences ranged from easy to difficult to lipread. The film offered no cues of context. The sentences were presented by a female speaker, showing the head, neck, and shoulders full face forward. The speaker appeared on the screen, made a statement, and the subject wrote down what he thought had been said. Exposure time was between two and four seconds for each sentence. Each word correctly recorded constituted one point towards the subject's total score. The maximum score possible for the 30 sentences was 188 points.

Measurement of Synthetic Ability. Synthetic ability was measured by scores on a letter prediction test. The test consisted of 20 sentences selected by a table of random numbers from the 30 sentences of the John Tracy Clinic Film Test of Lip Reading, Form B (7). For each sentence the subject was given a key word from that sentence. He was then to attempt to predict, one at a time, the letters of each word of the 20 sentences, the space between the words included. If the subject failed to predict a letter correctly on the first try, the examiner supplied the letter and the subject attempted to predict the succeeding letter. Thus, the element of synthetic ability is believed to have been shown in the subject's ability to anticipate the word from the knowledge of some of its letters and the sentence from the knowledge of some of its words. Each letter correctly predicted constituted 0.1 point toward the subject's total score. The maximum score possible for the 20 sentences was 56.8.

Experimental Procedure. The Film Test of Lip Reading was administered

to the 100 college students in groups of 10 or 12. The film was shown on a 16 mm projector in one of the classrooms in the Speech and Hearing Clinic. The subjects were seated in three rows of chairs with writing arms. The distance from the screen to the first row of chairs was 11 feet; from the screen to the second row of chairs, 13 feet; and from the screen to the third row of chairs, 15 feet. Because of the smallness of the group and the close proximity of its members, the distance variance among the subjects was not considered to be an important factor with respect to the test results. Prior to showing the film each subject was checked for an adequate view of the screen. The distance from the screen to the projector was 19.5 feet. The room was in semi-darkness, allowing sufficient light for the subjects to record their responses. Subjects wrote their names and their responses to the film test on a form sheet numbered from 1 to 30. Before the presentation of each sentence the projector was stopped and the subjects were alerted by the examiner's signal, 'Ready, watch the screen.' The subjects were given 25 seconds to record their responses for each sentence. The entire test took approximately 25 minutes.

The 50 experimental subjects were then given the vision and hearing screening tests. Any subject who failed the screening was to be eliminated from the experiment and to be replaced by the person with the next highest or lowest score as the case might be. In no instance, however, was it necessary to replace a subject.

Individual appointments were then made with the subjects for taking the letter prediction test. The subject was seated in front of the examiner and given the following instructions:

> This is a letter prediction test. I am going to give you one word from a sentence. You will then proceed to predict what is the first letter of the first word in the sentence. If your prediction is correct you will continue to the succeeding letter. If your prediction is wrong, I will supply the correct letter and you will attempt to predict the succeeding letter. The space between words should also be predicted. You will be allowed 15 seconds for each prediction. If you do not make a prediction within this time limit, I will supply the letter.

The examiner used a form sheet with the 20 sentences for scoring. The key word in each sentence was underlined. The examiner placed a one under each letter predicted correctly by the subject and a zero under each letter not predicted correctly.

Results

The null hypothesis under test in this study was that there is no difference between good and poor lipreaders with

TABLE 2. Means, standard deviations, and ranges of letter prediction scores for good and poor lipreaders, and t test for evaluating the difference between the two groups ($df = 48$).

Lipreaders	*M*	*SD*	*Range*	D_M	t
Good	49.8	.85	48-52	.5	1.80*
Poor	49.3	1.1	47-51		

*Not significant.

respect to synthetic ability as measured by letter prediction. The results of a *t* test showed the good lipreaders not to be significantly better letter predictors than the poor lipreaders. Table 2 presents a summary of the means, standard deviations, and *t*-test results.

The result of a point biserial correlation procedure, .27, although not tested for significance, indicated only a very weak, if any, relationship between the two sets of measures for lipreading ability and for synthetic ability (letter prediction ability). The evidence from the nonsignificant *t*-test result along with the low correlation coefficient thus strongly indicates that there is little relationship between lipreading ability and synthetic ability or, possibly, that letter prediction measures are not related to so-called synthetic ability.

Summary

This study was for the purpose of determining whether there is a relationship between lipreading ability, as measured by scores on a film test of lipreading, and synthetic ability, as measured by scores on a letter prediction test. Subjects were 50 male and female college students: 25 good lipreaders and 25 poor lipreaders as determined by film test scores. All subjects had normal hearing and vision.

The results provided no evidence of a difference between good and poor lipreaders with respect to letter prediction ability or of any important relationship between lipreading ability and letter prediction ability.

References

1. Bruhn, Martha Emma, *The Müller-Walle Method of Lip-Reading for the Deaf (Bruhn Lip-Reading System)*. Lynn, Mass.: Nicholas Press, 1929.
2. Howell, Louise, Lip-reading for the hard-of-hearing adult. *Volta Rev.*, 19, 1917, 15-16.
3. Mulder, R. L., A comparative study of the competence of groups of international and native students in aspects of language that hold relevance to speech. Ph.D. dissertation, Ohio State Univ., 1953.
4. Nitchie, E. B., *Lip-Reading Principles and Practice*. New York: Stokes, 1930.
5. Shannon, C. E., Prediction and entropy of printed English. *Bell Syst. tech. J.*, 30, 1951, 50-64.
6. Simmons, Audrey Ann, Factors related to lip reading. *J. Speech Hearing Res.*, 2, 1959, 340-352.
7. Taaffe, G., *A Film Test of Lip Reading*, No. II in *John Tracy Clinic Research Papers*. Los Angeles: John Tracy Clinic, 1957.

Shifts in Loudness of Pure Tones Associated with Contralateral Noise Stimulation

WILLIAM F. PRATHER

The term 'acoustic reflex' has been commonly used to denote the involuntary contraction of the two tympanic muscles, tensor tympani and stapedius, which occurs as a result of, among possibly other things, the exposure of an ear to loud sounds. Wever and Lawrence (*21*, chap. 10) and Kostelijk (*8*) have summarized considerable evidence that the middle ear muscles provide protection to the inner ear from intense acoustic stimuli and that the reflex is bilateral, that is, when one ear is stimulated by a sound adequate to result in contraction of the middle ear muscles, there is a contraction in both ears.

Evidence supporting the theory that the acoustic reflex significantly alters the transmission of sound through the middle ear can be found in studies of both animals and man. Procedures utilized in animal studies have included measurements of responses to sound stimuli by means of cochlear potentials when the tensor tympani has been severed (*2*, *22*), when tension was exerted on the stapedius or tensor tympani tendon (*5*, 7, *20*), and when spontaneous contractions occurred under dial urethane anethesia (*23*).

Studies in man have included changes in audiometric results by voluntary contractions of the muscles (*16*) and by an acoustic impedance matching bridge (*11*, *12*).

Although there is some diversity of opinion, the preponderance of evidence appears to indicate that the intra-aural muscles, especially the stapedius, give more protection for low frequency tones than for middle or high frequency tones, and that transmission of middle frequency tones may even be facilitated by contraction of the middle ear muscles (*14*, *17*, *20*, *23*).

Recently an attempt by Shapley (*14*) was made to quantify the amount of loudness reduction of pure tones thought to be caused, for the most part, by the contraction of the intra-aural muscles. In general, Shapley found loudness attenuation characteristics in his 32 female subjects similar to the energy transmission losses found by previous investigators in studies of animals. However, over-all attenuation values were substantially smaller than

William F. Prather (Ph.D., University of Iowa, 1960) is Instructor, Department of Speech Pathology and Audiology, and Department of Otolaryngology and Maxillofacial Surgery, University of Iowa. This article is based on a doctoral dissertation initiated under the direction of Dr. James L. Shapley and completed under the direction of Dr. Arnold M. Small, Jr.

those found for cats by other investigators. Also there were gross individual differences between subjects with regard to the amount of attenuation afforded by the acoustic reflex for any given set of experimental parameters.

One might conclude that these differences between subjects were real ones, that is, subjects actually do differ from one another with regard to the degree of protection that their middle ear muscles afford them. However, Shapley (*14*) also found that his within-subject reliability was low. Thus, inter-subject variability was confounded with intrasubject variability, and no clear conclusion could be drawn with respect to the intersubject differences.

According to Shapley,[1] the majority of his subjects reported voluntarily that pitch and quality changes occurred concomitantly with the loudness changes they observed. Pitch shifts in particular were said to have rendered loudness matching difficult. It would be interesting, therefore, to determine whether loudness measures, such as the ones that Shapley obtained, would be more stable if pitch shifts were in some way 'controlled,' and thus intrasubject variability reduced.

The primary purpose of this investigation was to measure the loudness change in pure tones as a result of noise stimulation in the opposite ear and to compare intrasubject and intersubject variabilities. These measures were obtained under two conditions: (a) the subjects were required to match for loudness without regard to any pitch changes, and (b) they were required to match for both loudness and pitch. It is, of course, implicitly assumed that the noise in the opposite ear (see *Procedure*) results in a binaural acoustic reflex and that this mechanism is responsible for most of the observed loudness and pitch changes.

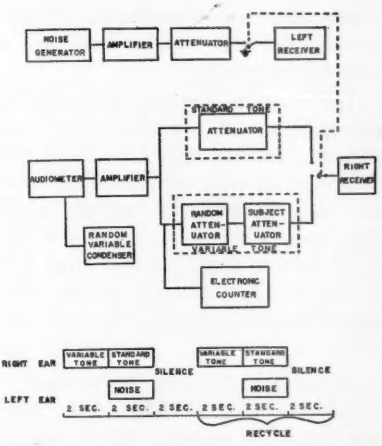

FIGURE 1. Schematic diagram of equipment and pattern of presentation of tone and noise stimuli.

Procedure

Figure 1 presents a schematic diagram of the experimental apparatus and pattern of presentation of the stimuli. The noise generator delivered a white noise to the receiver fitted in the left ear of the subject. The standard tone was unaltered except for frequency and intensity level adjustments by the experimenter. The frequency and intensity of the variable tone were adjusted by the subject. The experimenter also introduced small frequency changes and attenuation for the variable tone in a random fashion so as to

[1]Shapley, J. L., personal communication, 1959.

reduce effectively any visual and kinesthetic cues that might influence a subject's judgment.

The variable tone was presented to the right ear of the subject for two seconds. Immediately following this presentation, noise was introduced into the left ear for two seconds while the standard tone was presented in the right ear. A silent interval then followed for two seconds before there was a recycling of the two tones and noise.

Selection of Subjects. There were 10 subjects who were selected to meet the following criteria: (a) a hearing loss no greater than 10 db from normal hearing acuity for the frequencies 250 through 3 000 cps in each ear; (b) the ability to match within plus or minus 2 db the intensities of the variable (without noise) and standard tones; and (c) for the frequencies used in this study, the ability to match both the pitch and loudness of the variable (without noise) and standard tones to result in a pitch match within 2 DL for frequency, as described by Shower and Biddulph (*15*).

Subject Training. Preliminary procedures involving loudness matching and loudness-and-pitch matching were designed to serve the dual purpose of (a) selecting subjects according to the criteria described and of (b) training those selected in the techniques of matching. For the preliminary loudness balancing test, the noise channel was turned off and the subject was required to adjust the intensity of the variable tone so that its loudness equalled that of the standard. A 'stimulus schedule' was prepared in which each of the five frequencies used in the main study was presented at each of two sensation levels, 20 and 80 db. Each of the 10 combinations was presented twice, making 20 combinations in all, arranged in random order. For the preliminary loudness-and-pitch matching test, a similar procedure was employed. In the experimental condition which followed, the subjects thus chosen and trained were considered experienced listeners.

Earmolds. Impressions were made of each ear of each subject and subsequently used to fashion custom earmolds of the rigid, plastic, standard type.[2] These earmolds served as external canal plugs into which air-conduction receivers of the hearing aid type could be inserted to provide a good acoustic seal and to minimize cross hearing due to leakage of the sound stimuli around the head.

Conditions and Parameters. The following experimental parameters were used: (a) tone level: 20 db and 80 db sensation level; (b) noise level: 40 db and 100 db sensation level; (c) frequency: 250, 500, 1 000, 2 000, and 3 000 cps.

Each of the 20 combinations of the three parameters above was presented under the following two conditions: (a) loudness match alone; (b) loudness and pitch match.

Experimental Procedure. Each of the 20 combinations of the three parameters

[2]Impression material and earmolds were furnished by the Zenith Radio Corporation.

TABLE 1. Analysis of variance summary table for the loudness-match condition comparing tone level (TL) with noise level (NL) with frequency (F) for the criterion measure of average loudness change over five trials.

Source	*df*	*ms*	*F Ratio*	*F*
A (TL)	1	318.03	A/AS	5.54*
B (NL)	1	884.52	B/BS	25.94†
C (F)	4	35.71	C/CS	2.67*
S	9	128.76		
AB	1	11.11	AB/ABS	0.40
AC	4	20.14	AC/ACS	1.24
AS	9	55.36		
BC	4	97.98	BC/BCS	6.76†
BS	9	34.10		
CS	36	13.36		
ABC	4	41.60	ABC/ABCS	3.54*
ABS	9	27.87		
ACS	36	16.19		
BCS	36	14.50		
ABCS	36	11.73		
Total	199	31.34		

*Significant beyond 5% level.
†Significant beyond 1% level.

TABLE 2. Analysis of variance summary table for the loudness-and-pitch-match condition comparing tone level (TL) with noise level (NL) with frequency (F) for the criterion measure of average loudness change over five trials.

Source	*df*	*ms*	*F Ratio*	*F*
A (TL)	1	2540.42	A/AS	25.56*
B (NL)	1	410.12	B/BS	7.18†
C (F)	4	6.54	C/CS	1.01
S	9	232.31		
AB	1	2.17	AB/ABS	0.08
AC	4	16.45	AC/ACS	3.60†
AS	9	99.38		
BC	4	19.62	BC/BCS	2.11
BS	9	57.10		
CS	36	6.48		
ABC	4	17.60	ABC/ABCS	2.62
ABS	9	28.12		
ACS	36	4.57		
BCS	36	9.28		
ABCS	36	6.73		
Total	199	39.80		

*Significant beyond 1% level.
†Significant beyond 5% level.

was administered to each of the 10 subjects five times. These 100 combinations, then, were randomized for each of 10 lists, and the lists were randomly assigned to the 10 subjects. This procedure was used for each of the two conditions of the experiment. Subjects were alternated as to which condition was first.

For the loudness-balance-alone condition and for a given combination of parameters (predetermined according to the randomized list for a particular subject) each subject was instructed to match as best he could the loudness of the variable tone (no noise in the opposite ear) to the loudness of the standard tone (noise in the opposite ear). For a given combination of parameters recycling continued until the subject felt he had made the best possible match. The difference in level between the first and second tone, then, represented the shift in loudness brought about by the noise. This same procedure was followed for all combinations of parameters and for all subjects.

For the pitch-and-loudness match condition a similar procedure was followed except that the subject was instructed to match the two tones for both pitch and loudness by means of separate controls. The subject was allowed to match the two tones in any manner he chose, but, in general, each subject alternated between pitch and loudness matching, that is, he adjusted the tone first for pitch, then for loudness, or vice versa, and then back and forth between pitch and loudness adjustments until he felt he had the best possible match for *both*. In addition to recording the loudness shift values the experimenter recorded also the frequency values of the variable tones following the matches.

Each subject was allowed to stop a given listening session whenever he wished. Most subjects averaged approximately 50 minutes to one hour per listening session. Subjects varied in the number of listening sessions they needed to complete the experiment, depending upon the length of each listening session; however, in no case for any subject was the time greater than three weeks from the first listening session to the final listening session.

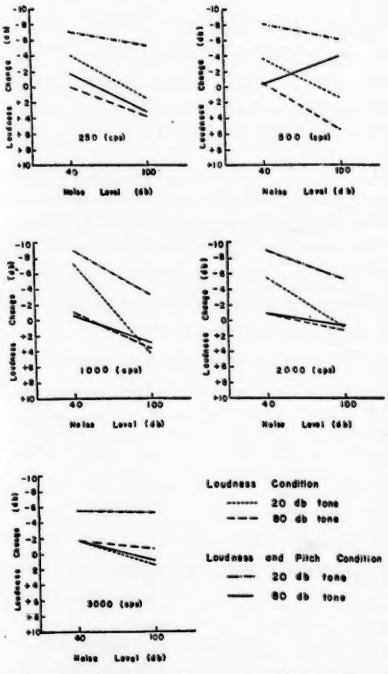

FIGURE 2. Loudness change required to match standard and variable tones for equal loudness, plotted as a function of noise level, tone level, frequency, and condition.

TABLE 3. Comparison of intrasubject variability between loudness-alone (LA) and loudness-and-pitch match (LP). This comparison is made for each of the 20 combinations of tone level (TL), noise level (NL), and frequency (F). The within mean squares are derived from the analyses of variance summarized in Tables 1 and 2.

Parameters			*LA*	*LP*	*F*
TL	*NL*	*F*	ms_w	ms_w	
20	40	250	9.48	17.40	1.84*
20	40	500	20.21	22.44	1.11
20	40	1000	18.16	30.52	1.68
20	40	2000	29.20	24.50	1.19
20	40	3000	41.00	30.68	1.34
80	40	250	4.32	8.08	1.87*
80	40	500	7.64	4.92	1.55
80	40	1000	6.04	5.84	1.03
80	40	2000	3.08	3.32	1.08
80	40	3000	4.32	5.76	1.33
20	100	250	16.52	36.40	2.20*
20	100	500	27.92	77.00	2.76*
20	100	1000	90.60	64.96	1.40
20	100	2000	82.04	61.76	1.33
20	100	3000	47.56	55.64	1.17
80	100	250	5.65	16.32	2.89*
80	100	500	8.84	10.60	1.20
80	100	1000	9.68	9.08	1.07
80	100	2000	6.76	5.32	1.27
80	100	3000	8.36	5.04	1.66

*Significant beyond 5% level, two-tailed test.

Results

The differences in level, specified in decibels, which were required to make the variable equal in loudness to the standard tone, were used as the values in the statistical analysis. These measures will hereafter be referred to as loudness changes. It is recognized, however, that since they are derived from a loudness equating technique, they are not loudness changes in the strict sense of the phrase.

Significance of Main Experimental Variables. The main experimental variables were evaluated by four-dimensional (*9*) analyses of variance, summarized in Tables 1 and 2, for each of the two conditions, (a) loudness match and (b) loudness and pitch match. The four dimensions are tone level (TL), noise level (NL), frequency (F), and subjects (S). The data for these analyses consisted of loudness changes averaged over five replications.

Presented graphically in Figure 2 are the mean differences in decibels required for equal loudness between the variable and standard tones for all trials and subjects.

Positive values on the loudness change axis indicate that the standard tone was reduced in loudness from its value when no noise was present in the opposite

ear. That is, with the noise and standard on simultaneously but in opposite ears, it was necessary for the subject to reduce the level of the variable tone in order that its loudness equal that of the standard. Negative values, on the other hand, indicate that a standard tone increased in loudness from its value when no noise was present in the opposite ear. That is, with the noise and standard on simultaneously but in opposite ears, it was necessary for the subject to increase the level of the variable in order that its loudness equal that of the standard.

The analysis of variance for the loudness-match condition shows that tone level, noise level, and frequency were factors significantly contributing to the total variance of the loudness-change values. The analysis of variance for the loudness-and-pitch-match condition shows that tone level and noise level, but not frequency, were significant factors. The graphs indicate that, in general, positive loudness changes (reduction in loudness of the standard tone) occurred for the 100 db SL noise, except for the 20 db tone level in the loudness-and-pitch condition. Negative values occurred for the 40 db noise levels and were more marked for the 20 db tone levels.

Comparison of Intrasubject Variability. Differences in intrasubject variability between the two conditions, (a) loudness match alone, and (b) loudness-and-pitch match, were evaluated for all 20 combinations of parameter values by means of a two-tailed *F* test, as reported in Table 3. *F*s were obtained by forming the ratio between the appropriate within-subjects mean squares. Since the mean squares forming each *F* ratio were related, the required assumption of independence was not met and it may be assumed that the test as applied is conservative.[3]

Table 3 indicates that five out of 20 *F* ratios are significant, and for each of these the within-subject variance was greater under the loudness and pitch match. In addition, it should be noted that four of these five significant *F*s occurred under the 250 cps frequency condition.

To obtain an index of consistency for five replications for all subjects, intraclass correlation coefficients were computed for each of the 20 combinations of parameters for loudness balance alone and for loudness-and-pitch match. The intraclass correlation coefficients were, in general, not large, and there did not appear to be any clear trend between conditions or between parameters.

Analysis of Variance for Frequency Changes. To evaluate frequency changes an analysis of variance, similar to that reported in Tables 1 and 2, was employed. Frequency change values in cycles per second were used instead of loudness change values. The frequency change represented the difference in frequency between the standard tone and the variable tone in the pitch-match condition. None of the sources of variance was significant.

Discussion

The results of the analyses of variance for loudness match alone and for

[3]That is, this application will increase the probability of failing to detect a difference when in fact a difference does exist.

loudness-and-pitch match, reported in Tables 1 and 2, agree in part with the results of Shapley (*14*) and Loeb and Riopelle (*10*). Shapley found a significant variance for noise level, although his noise levels ranged in 10 db steps from 60 db SL to 90 db SL, whereas in the present study only two noise levels were used, 40 db and 100 db SL. Shapley assumed that all four of his noise levels probably resulted in an acoustic reflex since all four levels produced an observable loudness change. Furthermore, he found an increasingly greater 'attenuation' of loudness as the noise levels were increased.

For the present study it was assumed that the acoustic reflex would be essentially inactive at the 40 db SL of noise. For no frequency or tone level at this level of noise did an apparent loudness reduction of the standard tone occur. Instead there appeared to be a facilitation effect which will be discussed shortly. On the contrary, however, for noise levels of 100 db with few exceptions the standard tone suffered a loudness decrease. This result is essentially in agreement with the results of Shapley (*14*) and Loeb and Riopelle (*10*).

Shapley's results, however, in contrast to those of the present study, provided no evidence of significant variance for tone levels, even though his tone levels ranged 40 db in 10 db steps, the lowest SL for each frequency being 40 db below the highest permissible SL. In the present study, on the other hand, only two levels, 20 db SL and 80 db SL, were employed. Had Shapley extended his tone level range to lower values, he might have found results similar to those of the present study and to the results of Loeb and Riopelle (*10*), that is, much less 'attenuation' of loudness for the tone levels nearer threshold than for the higher tone levels.

Loeb and Riopelle propose such a variation in 'attenuation' with tone level by means of a hypothetical model showing differential action of the ossicles and middle ear muscle tension for different amplitudes of the tone stimulus. They hypothesize that a wide excursion of the ossicles for the high tone levels is limited by the contracted and shortened middle ear muscles whereas the relatively smaller excursions for the low amplitude stimuli are less limited by the contracted muscles.

Such an explanation might also well apply to the results found in the present study. Such a model might be further extended to take into account the two different noise levels in the present experiment. The 40 db noise level resulted in no attenuation of loudness of the standard tone, and, in many instances, produced instead a facilitation of loudness. As to why there appears to be a slight attenuation of loudness for the 20 db SL tone at the 100 db NL in the loudness-alone condition and a facilitation of loudness in the loudness-and-pitch-match condition, the explanation is not immediately apparent.

Both Shapley (*14*) and Loeb and Riopelle (*10*) found a significant variance for frequency. The present study found an interaction with frequency for the loudness-match condition but not for the loudness-and-pitch-match condition. Again, an immediate explanation is not readily available.

The absolute attenuation values

(Figure 2) in the present study are small in comparison with those found by Wiggers (*23*), Galambos and Rupert (*4*), and Shapley (*14*). They are closer to those found by Loeb and Riopelle (*10*). Attenuation of loudness values in the present study ranged up to 5 db for the 80 db tone level in the lower frequencies. They appear to be slightly larger for the loudness-alone condition than for the loudness-and-pitch-match condition, although this difference is not marked. Differences in experimental procedures by various experimenters might account for such differences in the absolute values for the changes in loudness of the standard tone.

For 40 db of noise, facilitation of loudness was noted for all tone levels and frequencies for both matching conditions. For 100 db of noise such facilitation was noted for the 20 db tone level at all frequencies in the loudness-and-pitch-match condition. Facilitation has been noted by several investigators including Thurlow (*19*), Egan (*3*), Buytendijk and Meesters (*18*, pp. 1012-1013), and Rubin (*13*). Each of these investigators, however, has employed a somewhat different procedure from that used in the present study.

Rubin (*13*) has hypothesized that an attention factor might have at least partially contributed to his results. Thus the activating stimulus served to alert the subject that another stimulus was following, and he was more attentive to the latter than if no such warning had been given. In the present study it does not seem likely that such an explanation is applicable since the stimulus presentation pattern was recycled a number of times, thereby minimizing the 'alerting' function of the activating stimulus.

Apparently if the noise is of insufficient intensity to activate the middle ear muscle reflex(es), it may well interact centrally with the pure tone stimulus in the opposite ear in such a way as to increase the loudness of the pure tone. Further, if one is willing to accept the model of Loeb and Riopelle (*10*) as a reasonable one, a noise level of sufficient intensity to activate the middle ear muscles, such as the 100 db SL white noise used in this study, may result in a greater binaural summation of loudness, greater in magnitude than the attenuation resulting from muscle contraction. Thus the over-all net effect upon the loudness of the pure tone would be a facilitation effect, although perhaps not as great a facilitation as found in the 40 db SL noise conditions. The 40 db SL noise condition is one, of course, in which essentially no reflex activation would be expected and hence might be optimal for observation of loudness facilitation. Figure 2 appears to bear out this observation for the 20 db tone level in the loudness-and-pitch-match condition. Why in the loudness-alone condition similar results did not occur is not clear.

It was proposed at the onset of the present investigation that pitch shifts which sometimes appear (as reported by Shapley's[4] subjects) concomitantly with loudness changes may increase intrasubject variability when subjects attempt to make loudness matches under conditions such as the ones in this

[4]Shapley, J. L., personal communication, 1959.

experiment. It was hoped that by having subjects match for pitch as well as loudness, pitch would be, at least to a certain extent, controlled, and intrasubject variability would be reduced over a condition in which such pitch matches were not made. The results do not bear out this hypothesis. In fact, of the 20 combinations of parameters that were compared between the two conditions of loudness-and-pitch match and loudness-alone match with respect to intrasubject variabilities, the five significant *F*s were ones for which there was more variability in the loudness-and-pitch-match condition than in the loudness-alone condition.

Subjects also subjectively reported to the experimenter that, in general, they had much more difficulty in making the combination loudness-and-pitch matches than they had in making the loudness-alone match. Some reported that they concentrated so much on the pitch match itself that they were sometimes less concerned with the loudness matches in this same condition. Such a report might indicate that the criterion measure of loudness change would be more variable in the loudness-and-pitch-match condition unless the subjects were more highly trained than the ones used in the present experiment. Had the subjects been more highly trained, especially in making pitch matches, it is still conceivable that intrasubject variability might be significantly reduced in the pitch-match condition as compared to the no-pitch-match condition. A further explanation of these results, and perhaps a more plausible one, is that something other than pitch shifts (for example, duration of stimuli, interval between stimuli) accounts for the major part of the intrasubject variability.

It is entirely possible, of course, that cross-masking, either around or through the head, by the noise stimulus may have been at least partially responsible for some of the observed loudness changes. In order to estimate the possible effects of cross-masking with the present experiment a 1 000 cps tone of moderately high level was delivered to one ear. A second tone of about the same frequency was presented to the opposite ear and its level raised until monaural best beats occurred in that ear. By this method it was deduced that in the present study the intra-aural attenuation for 1 000 cps was approximately 56 db.

Zwislocki (*24*), with an acoustically shielded loud speaker connected to the ear by means of a plastic tube and secured in the bony meatus by means of a perforated ear warden, achieved 91 db interaural attenuation. There are at least two factors which may account for the discrepancy between this figure and the one obtained in the present study. Zwislocki went to great lengths to insure minimum transducer radiation, to provide a good seal, and to minimize bone surface to which the driving pressure is applied. The earmolds and hearing aid receivers undoubtedly fell short of the Zwislocki attainment in these respects. In addition, it has been shown by many investigators (*1*, *24*) that obturating the opposite ear enhances cross-hearing by means of bone conduction. Zwislocki closed only one ear, while in the present study both ears were fitted with earmolds.

It appears, then, that in the present study an attenuation between the two

ears of only 56 db may have resulted in some cross-masking effects. This would be especially pronounced for all combinations of 20 db tone with 100 db noise levels. Indeed, based on the data of Hawkins and Stevens (*6*), a threshold shift of nearly 5 db could be expected at 250 cps, 15 db at 500 cps, and 20 db at the other frequencies. How much loudness decrement is attributable to 'masking' in this experiment is not easily ascertained, since the acoustic reflex would presumably result in loudness changes in the same direction.

Summary and Conclusions

The loudness change in pure tones as a result of noise stimulation in the opposite ear was measured. The primary purpose was to compare intrasubject and intersubject variabilities of these loudness change values over five replications obtained for 10 subjects under two conditions: (a) matching for loudness without regard to any pitch changes and (b) matching for both pitch and loudness. The experimental parameters were tone level, noise level, and frequency.

Results indicated that tone level, noise level, and frequency in the loudness-match condition and tone level, noise level, but not frequency, in the loudness-and-pitch-match condition were factors significantly contributing to the total variance of the loudness change measures (difference in loudness between the standard and variable tones). Subjects were variable within themselves to about the same degree, irrespective of condition. One of the most important findings appeared to be that of a facilitation effect, that is, an apparent increase in loudness of tone under certain conditions of noise stimulation in the opposite ear.

Summario in Interlingua

Le cambio de altessa in tonos pur resultante de stimulation de strepito in le opposite aure esseva mesurava. Le proposito primari esseva comparar intrasubjecto e intersubjecto variabilitates de estes valutas de altessa cambio trans cinque repetitions obteneva pro dece subjectos sub duo conditions: (a) equante pro altessa sin regardo a alicun cambios de pitch (impression de frequentia sur le ascoltator) e (b) equante pro pitch e altessa. Le factors experimental esseva nivello de tono, nivello de strepito, e frequentia.

Resultos indicava que nivello de tono, nivello de strepito, e frequentia in le condition de equal altessa e nivello de tono, nivello de strepito, sed non frequentia, in le equal altessa-e-pitch condition esseva factors significantemente contribuente a le total variabilitate de le mesuras de altessa cambios (differentia in altessa inter le standard e variabile tonos). Subjectos esseva variabile intra se a approximativemente le equal extension, irrespective de condition. Un del le plus important trovantes apparite esser que de le effecto de facilitation, que es, un apparente augmento in altessa de tone sub certe conditions de stimulation de strepito in le opposite aura.

References

1. Bekesy, G. v., Vibration of the head in a sound field and its role in hearing by bone conduction. *J. acoust. Soc. Amer.*, 20, 1948, 749-760.

2. Crowe, S. J., Hughson, W., and Witting, E. G., Function of the tensor tympani muscle. *Arch. Otolaryng.*, 14, 1931, 575-580.

3. Egan, J. P., The effect of noise in one ear upon the loudness of speech in the other ear. *J. acoust. Soc. Amer.*, 20, 1948, 58-62.

4. Galambos, R., and Rupert, A., Action of the middle ear muscles in normal cats. *J. acoust. Soc. Amer.*, 31, 1959, 349-355.

5. Hallpike, C. S., On the function of the tympanic muscles. *J. Laryng.*, 50, 1935, 362-369.

6. Hawkins, J. E., Jr., and Stevens, S. S., The masking of pure tones and of speech by white noise. *J. acoust. Soc. Amer.*, 22, 1950, 6-13.

7. Hughson, W., and Witting, E. G., An objective study of auditory fatigue. *Acta Otolaryng.*, 21, 1934, 457-486.

8. Kostelijk, P. J., *Theories of Hearing.* Leiden: Univ. Pers Leiden, 1950.

9. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin, 1953.

10. Loeb, M., and Riopelle, A. J., Influence of loud contralateral stimulation on the threshold and perceived loudness of low-frequency tones. *J. acoust. Soc. Amer.*, 32, 1960, 602-610.

11. Metz, O., The acoustic impedance measured on normal and pathological ears. *Acta Otolaryng.*, Suppl. 63, 1946.

12. Metz, O., Studies on the contraction of the tympanic muscles as indicated by changes in the impedance of the ear. *Acta Otolaryng.*, 39, 1951, 397-405.

13. Rubin, H., Auditory facilitation following stimulation at low intensities. *J. acoust. Soc. Amer.*, 32, 1960, 670-681.

14. Shapley, J. L., A study of loudness reduction of tones presented to one ear during the exposure of the opposite ear to noise. Ph.D. dissertation, Univ. Iowa, 1954.

15. Shower, E. G., and Biddulph, R., Differential pitch sensitivity of the ear. *J. acoust. Soc. Amer.*, 3, 1931, 275-287.

16. Smith, H. D., Audiometric effects of voluntary contraction of the tensor tympani muscles. *Arch. Otolaryng.*, 38, 1943, 369-372.

17. Smith, K. R., and Wever, E. G., The problem of stimulation deafness. III. The functional and histological effects of a high-frequency stimulus. *J. exp. Psychol.*, 39, 1949, 238-241.

18. Stevens, S. S. (Ed.), *Handbook of Experimental Psychology.* New York: Wiley, 1951.

19. Thurlow, W. R., Studies in auditory theory. I. Binaural interaction and the perception of pitch. *J. exp. Psychol.*, 32, 1943, 17-36.

20. Wever, E. G., and Bray, C. W., The stapedius muscle in relation to sound conduction. *J. exp. Psychol.*, 31, 1942, 35-43.

21. Wever, E. G., and Lawrence, Merle, *Physiological Acoustics.* Princeton: Princeton Univ. Press, 1954.

22. Wever, E. G., and Vernon, J. A., The effects of the tympanic muscle reflexes upon sound transmission. *Acta Otolaryng.*, 45, 1955, 433-439.

23. Wiggers, H. C., The functions of the intra-aural muscles. *Amer. J. Physiol.*, 120, 1937, 771-780.

24. Zwislocki, J., Acoustic attenuation between the ears. *J. acoust. Soc. Amer.*, 25, 1953, 752-759.

► *RESEARCH NEWS NOTE*

Special Kind of Speech Noise

►The speech science laboratory of the School of Dramatic Art and Speech at Ohio University, Athens, has recently completed the development of a special kind of speech noise. The noise is composed of the speech of human speakers and its spectrum is flat across the speech range. Research now in progress aims to test the masking efficacy of this noise in audiological and intelligibility studies, in comparison to saw-tooth noise, white noise, or filtered white noise. Inquiries should be addressed to Dr. Edward M. Penson.

A. C. LaFollette, Ph.D.
Director, Speech and Hearing Center
Ohio University, Athens

Letters to the Editor

The Editorial Staff assumes no responsibility for the opinions expressed in letters.

Suggests Relation of Linguistic Theory, Speech Therapy Not Appreciated

During the past few decades there have been a number of futile attempts to demonstrate to the satisfaction of practical speech therapists the intimate relationships between (1) linguistics and speech therapy, and (2) experimental phonetics (or speech science) and speech therapy. In 1957, for example, the annual convention of ASHA attempted (rather unsuccessfully from my point of view) to demonstrate and strengthen these relationships by providing sectional meetings entitled 'Speech from the viewpoint of linguistic science' and 'Current research on examination and diagnosis of articulation from the speech scientist's viewpoint.' In the same vein, the recent report, 'Research needs in speech pathology and audiology,' contains a statement by the subcommittee on problems of basic research in speech and hearing which assigns basic problems to two areas—language and speech production. Furthermore, the introduction to the latter area asserts that 'An understanding of *linguistic structures* in terms of the mechanisms of speech production is particularly needed' (my italics).

In the light of this situation the publication in the September 1959 issue of *JSHR* of an article entitled 'A phonetic study of misarticulation of /r/' seems to represent a potential landmark in the literature of speech therapy. The 'potential' is used advisedly; my responses to the article in question are ambivalent and require some analysis and clarification.

At the outset let me make it clear that I recognize the need for the research reported in the article. I also believe that a respectable experiment was performed. I am concerned primarily with what seems to me to be (1) misleading interpretations of the data, (2) a failure to fully exploit the data, and (3) a failure to relate the experimental problem to some basic issues of a pedagogical nature. These problems, in part at least, may lie outside of the interest and concern of the authors of the article in question, and, therefore, my remarks need not be interpreted as a censure of them.

The discussion that follows will try to develop the antitheses that Curtis and Hardy have both (1) demonstrated the relations between linguistic (i.e., phonemic) theory and speech therapy, and (2) failed to appreciate these relations. Furthermore, it will emphasize that an opportunity to point out a glaring deficiency in the armamentarium and/or training of speech therapists has been overlooked.

1. In the title of the article the symbolic expression '/r/' is used. The various contexts in which this and similar (i.e., /s/) expressions occur suggest strongly that a symbol enclosed in this fashion represents a *phoneme*. Furthermore, in various places in the text, symbols are enclosed in brackets (viz., [r]) to indicate some kind of phonetic representation. Note, however, that when one writes

> If, . . . , the /ɪ/ in the [tɝkɪ] was perceived to contain elements of both the /ɪ/ and /i/, the /i/ symbol was placed just above the /ɪ/ symbol . . . (p. 247)

or,

> . . . the symbol [r] to represent the consonant glide /r/, [ɝ] to represent the stressed vocalic /r/, and [ɚ] to represent the unstressed vocalic /r/ (p. 247)

assumptions contrary to general linguistic theory are being made about the nature of the phoneme.

The assumption that seems to be operating here is that phonemes are describable or definable in terms of articulatory events. My belief that such an assumption is implicit in the article is borne out quite clearly in the Discussion (p. 255), where the authors conclude that, since their study demonstrates the existence of various kinds of 'phonetic events to which subjects show differential behavior,'

grouping these events into a general classification for clinical purposes may be unwise, *and suggest furthermore* that underlying differences in articulation 'be considered in relation to the problem of their appropriate phonemic classification.'

This assumption by the authors is internal evidence of a failure to amalgamate their subject with modern linguistics. Contemporary phonemicists or phonologists are not apt to spend time correlating articulatory events and phonemic entities in a one-to-one fashion. On the contrary, the most vigorous streams of linguistic thought describe phonemes as contrastive elements without reference to their articulatory manifestations. It is interesting to point out that in concluding a discussion of *r* in various languages, Trubetskoy wrote (Principes de phonologie, p. 76; my translation of Cantineau):

> One ought not fool oneself by using identical signs of transcription: these signs are only an expedient. If one can only designate by the same letter the phonemes having a completely identical phonologic content, one would have to use a unique alphabet for each language.

The phonemic-phonetic problem has some interesting ramifications. For example, Curtis and Hardy state (p. 255) that their data suggest 'that correct production of consonant sounds usually misarticulated as singles may be facilitated in blends.' This statement, of course, must be read as a statement about phonemes in the distributional sense (and not articulatory entities). On the other hand, if my assumption is not correct, there seems to be reason for comparisons to be made either between different phonetic events (i.e., phonologically-distributed allophones) or between different 'articulatory phonemes.'

If the above analysis is indeed true, then this problem can be extended still further by asking whether a phonemic analysis of the dialect spoken by the children of the study shows allophones of /ɹ/ describable as [ɝ] and [ɚ], and whether the *r*-coloring in vowels and offglides is mandatory (i.e., phonemic) in certain phonetic environments. This point bears heavily on the interpretation of the responses alleged to be errors of the type [ɝ] and [ɚ].

2. It is worth mentioning here that one of the most interesting outcomes of the analysis made by Curtis and Hardy corresponds in a sense to a suggestion for the use of symbols made some time ago by Hultzén. In his essay on phonetic methodology alluded to by the authors (their ref. 5), Hultzén discusses a systematic procedure for selecting symbols for transcription; he points out emphatically, however, that his procedure is not based on any phonemic considerations. My interest in the relation between Hultzén's ideas and their application by Curtis and Hardy is intensified by the fact that for many years Hultzén has attempted to point out to speech-oriented people the advantages that linguistic know-how has for them. In particular, attention might be called to his articles entitled 'Phonetics, phonemics, and teachers of speech' (Quart. J. Speech, 33, 1947, 202-206) and 'Phoneme, allophone–segment' (Quart. J. Speech, 40, 1954, 260-268) in which the relations between phonetics and phonemics are delineated clearly and where their joint application to problems of speech teaching and rehabilitation are discussed.

The importance of linguistic analysis is readily apparent if we can bring ourselves to understand that everything such an analysis put into the experiment under discussion came out in the results. Such a state of affairs can be interpreted to mean that the experiment succeeded in identifying something, but it can also be said that the elements previously identified by different procedures are indeed present, i.e., as a validation of the original assumptions. In the present case I would advance the point of view that the Curtis and Hardy experiment is interesting to linguists because it provides confirmation of linguistic concepts by means of a new technique–an error count–and is interesting to speech therapists because it demonstrates how linguistic concepts can lead to fruitful insights about speech production.

On the other hand, this experiment is not a revelation about the articulatory effects of phonetic environment, since these effects have been described and discussed rather well in the linguistic literature. Similarly, it casts no new light on problems of phonemic classification, etc., since it does not utilize appropriate techniques to grapple with such subjects.

In passing it should be pointed out that treating *r* as a consonant or as a vowel, etc., should not surprise anyone. Almost any standard phonetic work provides many examples of these interesting problems of articulation and/or classification. Heffner's monograph (*General phonetics*, 1952), for example, discusses *r*-sounds as intermittent stops ([ɾ], [ɽ], [ʀ]), open fricatives ([ɻ], [ɽ], [ʁ], [$^{\text{R}}_{\omega}$]), and reduced *r*-forms (including such things as [ɝ], [ɚ]); even Jones (*An*

outline of English phonetics, 8th ed., 1956) provides an appendix on American English in which [ɹ] is listed as a vowel in *bird* and *water.*

A very real problem, of course, is always with us—the pedagogical one. I have the uncomfortable feeling that most 'speech people' pick up their ideas about phonetics and phonemics in highly oversimplified accounts (usually found in speech texts). The danger of this procedure has been noted before by Hultzén (1947), who said that

> . . . we have permitted ourselves to be confused by unconscious phonemic simplifications. For example, a great many elementary works on phonetics classify consonants as voiced or voiceless without paying any attention to fortis or lenis characteristics, i.e., the force of breath and vigor of articulation, or of aspiration. It is really a phonemic procedure, if a bad one, that we simplified classification to what we have thought were essential details, putting aside what we considered accidental features.

3. There are a number of points about the treatment and interpretation of data from the experiment that are puzzling. In Table 3, for example, we find the entry [ŋɚ], a syllable which many would suppose does not occur in English; an example would be helpful here.

In a similarly trivial vein, in Table 5 no errors of omission are noted for [ɚ]. This probably means that an assumption was made that [ɚ] was *one* sound and not divisible into [ɜr], for example, and therefore, an error would result in a substitution (or a distortion), but not an omission. On the next three lines of the table, however, there are entries both under *omission* and [ə], meaning that [ɚ] was divisible into [ər]. I wonder why.

On a more serious note, however, I am confused by the use of the word *systematic* in descriptions of inconsistency of articulation. My confusion exists on two levels—the trivial problem of how something can be inconsistent in a systematic way I will not dwell upon. I am concerned, however, with the rationale used to establish the thesis that the present data reflect systematic tendencies of 'correct productions of [r] associated with place of articulation' (p. 250). The evidence here is not in any sense unequivocal; certainly a more comprehensive explanation is suggested by the data.

The intepretation of the data shown in Table 7 (where the source of comparison data is incorrectly stated) is labored enough to weaken the general argument. Examination of the tabulated data does not seem to agree with the first two of the three assertions made by the authors (pp. 253 ff).

4. There are interpretations of a more basic nature that trouble me. The last paragraph on p. 252, for example, sums up the discussion of 'Differentiation among types of /r/' by asserting that the various differential responses should be considered in the analysis of a child's misarticulations. This is a healthy point of view, but let us continue the discussion. The authors now state that 'the evidence argues convincingly against classifying all of these /r/ sounds within a single category implying relationship with respect to articulatory movements.' Here is where my concern begins. I wonder who has been guilty of such misbehavior. Certainly not the phonetician—he knows better. Certainly not the phonemicist—he knows better. Certainly not the teacher of speech correction—he should know better!

I suspect that the statement is either an attempt to set up a straw man, or is a reaction to the state of linguistic naïvete in the field of speech therapy. The straw man is easily overthrown, but more convincingly by the corpus of linguistic knowledge than by the results of this particular experiment.

5. Why then is this study important? If for no other reason, this study is important for its contribution to the literature on 'articulation testing.' While it stresses ideas that I am sure others have used in their classes—I have done so myself—these ideas are not discussed adequately in most therapy texts. The following points need emphasis:

a. The tester of articulation should present materials organized on the basis of syllables rather than words; i.e., it is more fruitful to consider that consonant sounds appear initially or finally in syllables rather than initially, medially, and finally in words.

b. The tester of articulation should transcribe responses in symbols that describe articulatory events whenever possible; i.e., a narrow phonetic transcription is a valuable tool for the therapist.

c. By the proper exploitation of phonetic skills, the results of articulation tests traditionally called *distortions* will be minimized (as a function of the transcriber's phonetic sophistication) and the results called *substitutions* will be maximized.

d. Articulation testing materials may require organization into two levels—gross

screening materials, and subtests for the study of the allophones of particular phonemes.

Finally, the article is important for having provided an opportunity for clarifying the relations between linguistic skills and practical speech therapy.

Arthur S. House
Research Laboratory of Electronics*
Massachusetts Institute of Technology

*Work of this laboratory supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research.

Cite Importance to Therapy of Regarding Phoneme as Articulatory Event

Thank you very much for allowing us the opportunity to reply to the letter of Dr. Arthur House.

We appreciate very much both the complimentary things which Dr. House was kind enough to say concerning our article and also the compliment implied by the fact that he has thought our article deserving of serious criticism. Possibly because we are at least partially guilty of 'naïveté' in linguistic theory as charged by Dr. House, we had not considered our article as a 'potential landmark in the literature of speech therapy.'

As we understand Dr. House's letter, his several points may be summed up briefly as follows: (1) He considers us to have been somewhat naïve concerning general linguistic theory and to have made certain errors in consequence thereof. (2) He does not agree with certain of our interpretations of the data. (3) He believes we should have gone much farther than we chose to go in developing certain philosophical points, especially concerning the relations of linguistic theory to speech correction and what he terms 'a glaring deficiency in the armamentarium and/or training of speech therapists.'

Concerning point number one in the above list, we readily plead guilty to knowing less about the field of linguistic theory than we would like to know. We would as readily plead guilty to knowing less in the fields of acoustics, physiology, psychology, statistics, electronics, etc., than we would like to know, or than we ideally should know to be most effective in our day-to-day work. In a field which crosses as many disciplines as ours, we imagine these feelings are shared by many colleagues. If our lack of expertness as linguists resulted in errors, we apologize, and we are grateful to Dr. House for making such corrections as will make the presentation more accurate and effective.

Dr. House has criticized our use of brackets and slash marks, as well as one particular statement in our discussion, because, in his words, '. . . assumptions contrary to general linguistic theory are being made about the nature of the phoneme.' The objectionable assumption which he ascribes to us is that '. . . phonemes are describable or definable in terms of articulatory events.' The authors are certainly aware that the integrity of a phoneme depends primarily on its contrastive significance in the language, rather than its description as a highly specific articulatory event. If we gave a contrary impression, we communicated poorly and Dr. House would seem to have grounds for criticism. We will freely grant that the statement referred to on p. 255 was probably unwise, and could have been omitted without weakening the general import of the article.

However, we find some difficulty in accepting without reservation Dr. House's statement that, '. . . the most vigorous streams of linguistic thought describe phonemes as contrastive elements *without reference to their articulatory manifestations*' (italics ours). This statement implies that articulatory considerations have no significance whatever in determining linguistic contrasts, a position that the authors would not wish to try to defend. Most writers in general linguistic theory make frequent reference to articulatory features in their discussions of phonemes, and such features are normally utilized as one of the principal bases of classification. A statement such as that quoted below from Hockett, *A Course in Modern Linguistics*, p. 24, might be interpreted as supporting Dr. House's position:

> 'The phonological system of a language is therefore not so much a "set of sounds" as it is a *network of differences between sounds*.'

However, one can hardly escape the further question, which is left completely untouched by this statement, concerning the nature of the differences which comprise the network. Although the answer to this question is certainly far from simple, and must include such complex factors as patterns of learned similarities and differences, it is difficult to imagine any complete account of the nature of the

network of differences which does not consider the acoustic differences to which the listener reacts, or the articulatory events by means of which these acoustic differences are generated. The writers' admittedly limited reading of linguistic literature would not lead them to the conclusion that these factors are actually ignored, although they may be assigned a secondary role by the phonemicist whose principal problem is to identify the network of sound contrasts which will permit him to describe a language.

For the speech correctionist, on the other hand, the acoustic and articulatory variations that contribute to the 'network of differences' are part and parcel of his primary problem. He is necessarily concerned with a description of articulatory behavior not only in relation to the sound contrasts which are significant in a language, or a dialect, but also in relation to a standard of acceptability or correctness for 'normal' articulatory behavior. Such description necessarily and frequently requires discriminations which are not involved in linguistic contrast. In this connection, where the authors would seem to have gotten into difficulty is in attempting to adapt to their purposes a system of notation which was not designed for their kind of descriptive problem. That this has resulted in something less than perfect understanding is probably not surprising. The authors were highly aware of difficulties in using this notation while the article was in preparation, because we could foresee difficulties both in communicating our meaning and in implying more than we intended. While we regret that we did not succeed better in avoiding these difficulties, we do not believe the solution is quite so simple as Dr. House's remarks suggest.

Considerable space would be required to deal with each of the detailed points which Dr. House has raised concerning interpretations of our data. Certainly we would accord anyone the privilege of disagreeing with our interpretations. However, we would like to consider further a few of the questions he has raised. To begin with, Dr. House does not seem altogether consistent in his criticism. At one point he states, '. . . this experiment is not a revelation about the articulatory effects of phonetic environment, since these effects have been described rather well in the linguistic literature.' In another place, however, he seems to deny the validity of our interpretation that such effects exist when he questions the basis for our statement that relationship has been shown between correct articulations of /r/ and certain features of phonetic context. The evidence may not be 'unequivocal,' to use Dr. House's term, but the rank-order of the precentage of correct /r/ responses as shown in Table 7 is very nearly the rank-order that would be predicted, if one were to assume the relationship stated in our interpretation. In his second point concerning our Table 7, Dr. House seems not to have understood that our statement was based on a comparison of the data from Table 3 with the data of Table. 7. If the appropriate comparisons are considered, we believe that our statement is clearly supported.

We would also choose to disagree that the effects described in the experiment 'have been described rather well in the linguistic literature.' Certainly the general facts of interaction between sounds which are contiguous in time have been previously described in numerous studies. But our point had to do with a particular case, viz., the specific influence of phonetic context in facilitating correct articulations of sounds in the speech of children who do not so speak as a matter of established habit. While one might have predicted the outcome from what was known about the influence of phonetic context on articulation, it does not follow that this prediction is entirely obvious or the only possible result in this special case. To the authors at least, the empirical test has therefore seemed worthwhile.

Throughout Dr. House's letter it seems apparent to us that he was reading our article from a quite different point of view than that from which we wrote. Thus, his letter emphasizes the linguistic implications, as well as faults, of the study. To us the primary purpose was to describe the articulatory behavior of children with respect to the /r/ sounds in such a way that speech correctionists could be helped to test articulatory behavior more meaningfully and plan corrective procedures more effectively. Dr. House has been kind enough to state that the study has indeed made a contribution to the literature on articulation testing. This was our primary intention.

In this context it seemed relevant to make a point of classifying vowel /r/s and consonant /r/s separately. Indeed, such separate classification was an integral part of the analysis, although we made no claim to having discovered anything new as Dr. House implies. However, the point which we tried to make, viz., that the articulatory behavior of children varies systematically with respect to these subclassifications of /r/, both with

respect to frequency of correct articulations and the nature of errors that occur, has not been previously developed. Moreover, it could not be developed without data of the kind obtained in our study.

Lastly, in this context, we do not believe that we have been guilty of 'setting up a straw man.' There is ample evidence in the literature of articulation testing that the various subclasses of /r/ are often considered as comprising a single articulatory entity, so far as testing articulatory correctness and instituting corrective procedures are concerned. Probably, as Dr. House states, this should not happen. The fact is that it does. Granted that linguists have long made distinctions between a number of allophones of /r/, it still had not been demonstrated that these were distinctions which make a difference in evaluating the articulatory behavior of children, or planning a corrective program to correct articulatory errors. We hope that our joust with the 'straw man' may have provided this demonstration.

In conclusion, we feel that the disagreement between Dr. House and us is not as great as his letter and our response would make it seem. As previously stated, we think a large part of the difference stems from the fact that he read from a different point of view than that from which we wrote. In part at least, he read and criticized our article as an investigation in linguistics. We wrote with a much more limited purpose. Had we written with Dr. House's orientation, we might have undertaken the broader task that he wished us to tackle. Because we had a more limited and specific purpose, we did not feel that it was appropriate for us to do so. We certainly agree with him that a greater understanding of the principles of linguistics would be helpful to persons interested in speech disorders, and we hope that gains in this direction can be made.

James F. Curtis, Head
Department of Speech Pathology
and Audiology
University of Iowa
James C. Hardy, Supervisor
Speech and Hearing Services
Hospital School for Severely
Handicapped Children
University of Iowa